

OY	2273	ATGCCCATTCGCCACGAGTTCTCACTGATGTRGTC--CTAGTCTCCGGCGGTTTAT	2322
Db	540	MetProIlealIahISgluPheSerProaspValValTrhLeuValSerAlglYpHeasp	559
OY	2330	GCTGTGAGACATCTGTCTCCTCGGGGGCTACTTCGTACCGCCAGATGTTTGCC	2389
Db	560	AlaValIgluGlyHisLeuSerProLeuGlyGlyTyrSerValTrhAlaIArgGlyPheGly	579
OY	2390	CACTTGACACGAGCAGTATGACCCCTGGCAGGCGGGCGGGTGCTGGCCCTGGAGGA	2449
Db	580	HisLeuThrArgIuLeuMetTrhLeuAlaGlyGlyArgValValLeuAlaLeuIuGly	599
OY	2450	GGCCATACATTGACGCCCATCTGTATGCCCTGAAAGCTTGCTGGCTCTGCTACT	2509
Db	600	GlyHisAspLeuThrAlaIleCysAspAlaSerGluAlaCysValSerAlaLeuLeuSer	619
OY	2510	GTTAAGGTGGACGCCCTGGATGTAGGGAAGTCTGGACCAAAACCAATCAAGCACTG	2569
Db	620	ValIuLeuGlnProLeuAspGluLeuValLeuGlnGlnLysProAsnIleAsnAlaVal	639
OY	2570	GCCACGCTAGAGAAAGTCATCAGAGATCCAGAGCAAAACATGGAGCTGTGCAGCAATTC	2629
Db	640	AlaThrIleuGluIuValIleGluTrhGlnSerLysHisTrpSerCysValGlnLysShe	659
OY	2630	GCCGCTGCTGTGGCGCGGTCCCTGGAGAGGGGCCCAAGCAGGTAGACCGAAGAACCCGA	2689
Db	660	AlaIaIaGlyLeuGlyArgSerLeuArgIuAlaGlnAlaGlyGlyIuTrhGluIuIaGlu	679
OY	2690	AT-GTGAACGCCATGGCCTTGCTGTGGTGGGGGGCCGAACAGGCCCAAGCTTGGCAACC	2748
Db	680	ThrValSerAlaMetAlaLeuLeuSerValGlyAlaGlnGlnAlaGlnAlaAlaIaIa	699
OY	2749	CGGGAACACAGCCCGACCGCGGAGAGAGGCCATGACGACAGACCTGGCCCTG	2802
Db	700	ArgGluHisSerProAlaGlyIuGluIuProMetGluGlnGluIuProAlaLeu	717

```

US-10-072-094-7
? Sequence 7, Application US/10072094
? Publication No. US2003002538A1
? GENERAL INFORMATION:
? APPLICANT: JACKSON, DONALD
? APPLICANT: LORENZINI, MATTHEW
? APPLICANT: ATTARI, RICARDO
? APPLICANT: GOTTFARDS, MARCO
? TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
? FILE REFERENCE: 3053-414505U1
? CURRENT APPLICATION NUMBER: US/10/072,094
? CURRENT FILING DATE: 2002-06-14
? PRIOR APPLICATION NUMBER: 60/298,296
? PRIOR FILING DATE: 2001-06-14
? NUMBER OF SEQ ID NOS: 127
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 7
? LENGTH: 1084
? TYPE: PRT
? ORGANISM: Homo sapiens
US-10-072-094-7

Alignment Scores:
Pred. No.:      8,15e+15
Score:          2691.00
Percent Similarity: 71.38%
Best Local Similarity: 59.11%
Query Match:     50.34%
DB:              9

US-09-502-945-2 (1-2885) x US-10-072-094-7 (1-1084)

Length:      1084
Matches:     568
Conservative: 118
Mismatches:  195
Indels:      61
Gaps:        23

```

QY 2 GAATTCCTCTTGTGCAAGTCAAGAGCCACACAGCGGCTCAACCATTCCTCCCA 61
 |||||:::||||: ||| |||||:::
 Db 177 G|upheValLeuAsnLysLysLysAlaLeuAlaHisArgAsnLeuAsnHisCysTieSer 196

OY	62	CAGACACCCCAATCTGCG-----GAGGCCACCAATGCTTCTTGGACACAGACTGCCCT	115
Db	197	SerAspProValGlyTyrTyrGlyLysThrGlnHisSerSerLeuAspGlnSerPro	216
OY	116	CCCCAGACGGGCGCCCGGGAGCGCCCTCCCTCCACAAATGCGCTTGGCTGGGGCCCTAC	175
Db	217	ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetIrr	233
OY	176	GACAGCTGAGACGACTTCCCTCCCTCCGAAAACAGCTCTGTAAACCCAACTTGAAGTGCCT	235
Db	234	AspAlaLysAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysLeuArg	253
OY	236	TCAAGGCTAAACAGAGAGTGGCTGAGCGGAGAGACAGTCCCTCTGGTGGCAAGAT	295
Db	254	SerArgLeuLysGlnLysValAlaGlnArgSerSerProLeuLeuArgTyrLysAsp	273
OY	296	GGAGCTGTTATTACACCTTTAAAGACAGAGCGTTGAGTACAGAGTGGCGGGGCTGGG	355
Db	274	GlyProValValThrAlaLeuLysLysArgProLeuAspValThr-----	288
OY	356	GCGTCGTCCTGTGTAAACAGCGACCGCGGCTCGGGCCAGCTCTCC---AACAGCTCC	412
Db	289	---AspSerAlaLysSerSerAlaProGluSerGlyProSerSerProAsnAsnSer	307
OY	413	CACAGACCAACGCTGAGAAATGGCTTTACTGGCTCACTGCCCAACATCCCCACTAGATG	472
Db	308	GlySerValSerAlaGlnAsnGlyIleAlaProAlaValProSerIleProAlaGluThr	327
OY	473	CTCCCTCAGCAGCGAGCGCCCTCCCTGTGACAGCTCCCGCCCAACCAAGTTACAGCCTACAG	532
Db	328	SerLeuAlaHisArgLeuValAlaArgGlnGlySerAlaAlaProLeuProLeuTyrThr	347
OY	533	TCTCCTCTCTCGCCCAACATCTCCCTGAGGCTCGAGCGCACGGCTCACTGTCCAACTCA	592
Db	348	SerProSerLeuProAsnIleThrLeuGluLeuProAlaThr-----	361
OY	593	CACCTCACTGCTCCCGAGACGTGTGCACA-----CAGCAGAGGCCCGAGAGCGAGCC	646
Db	362	-----GlyProSerAlaGluTyrAlaGluGlnIleAsnAspArgGlnArgLeuThr	377
OY	647	CTCCAGTCCCGCGGGGAC-----GCTGGACGCTGTACACGGCAAGTTCAAG	691
Db	378	LeuProAlaLeuGlnGlnAlaArgLeuSerLeuPheProGluThrHisLeuThrProTyrLeu	397
OY	692	AGCAGACTCTTATTCTCTGGGCTGCTGAGCGTGAGCTGAGGAGCGGACGGACCC	751
Db	398	SerThrSer-----ProLeuGlnArgAspLeuArgAla	408
OY	752	CACGGGACATGCTCCCTGTGCAGACATAGTCTGTGCTGAGCAGGCGCCGGACAGAC	811
Db	409	---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGluGlnProProAlaGlnAla	427
OY	812	ACCTGCAAT-----GCTGTGSCCACTCCAGGGGACAGTCCCACTAGTGACGGT	859
Db	428	ProLeuValThrGlyLeuGlnAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla	446
OY	860	GAACTGTGGCCACCAACATGCGGAGCGTAGAGAACTCCCGGAGATGGCCCGCTTAC	919
Db	447	AspArgValSerProSer-----IleHisLysLeuArgGlnHisArgProLeuGly	463
OY	920	CGCACTAGTCTCTACCGCTGCGCGACAGATGSCCAGGCGCTGTGACAGCGCTGTACGA	979
Db	464	ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln	483
OY	980	CACAAGCAGCAGAGTTCTGTGAGAGCAGAG-----CAGCAGACGCTTACG	1027
Db	484	GlnGlnHisGlnGlnProLeuGlnLysLysLeuGlnGlnProGlnGlnGlnGlnLeuGln	503
OY	1028	CTGGGCAAGATCTCCACCAACAGAGGAGGACTGCCACGAGCCCAACCCACCTTACG	1087
Db	504	MetAsnLysIleIleProLysProSerGluProAlaArgGlnProGluSerHisProGlu	523

```
QY 1028 CTGGGCAAGATCCCTACACACAGGGAGCTGCCCCAGGACCCACACCCACCTGAG 1087
      :: |||||::: || || |||||::: |||||
Db 504 MetasnlYslelleProlYsProserGluProlaHrgLmProGluSerHIsProGlu 523
```


Best Local Similarity: 59.00% Mismatches: 196
 Query Match: 50.22% Indels: 81
 DB: 10 Gaps: 23

US-09-502-945-2 (1-2885) x US-09-817-913-7 (1-967)

QY 2 GAATTCCTGTGTCGAAGTCAAGAGCCACACAGCGCGCTCAACATTCCTCCCA 61
 Db 60 GlnbValLeuAnslLysLysAlaLeuAlaHisProAsnLeuAnslLysCylLser 79

QY 62 CAGACCCCAATCTGG-----GGAGCCACCATGCTTCTTGGACAGAGTCCCT 115
 Db 80 SerCysProArGlyTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerSerPro 99

QY 116 CCCAGAGCGGGCCCCCTGGAGCCCTCCCTCCACAACTGCTTGGCCCTGGCCCTAC 175
 Db 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetCyl 116

QY 176 GACAGTCGAGACGACTTCCCTCCGAAAACAGCTCTGAAACCACTTGAAGTGGCT 235
 Db 117 AspaLalysAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysLeuArg 136

QY 236 TCAAGGCTAAACAGAGGTGGCTGAGCGGAGAGAGTCCCTCTGCGCAGAGAT 295
 Db 137 SerArgLeuLysGlnLysValAlaGluArgSerSerProLeuLeuArgLysAsp 156

QY 296 GGGACTTATTAAGACCTTTAAGAGAGACTTGAAGTACAGAGTCCGGGCTGG 355
 Db 157 GlyProValAlaThrAlaLeuLysLysArgProLeuAspValThr----- 171

QY 356 GCGCTGCTGCTGTACAGCGCAGCGCGCTCCGCGCCAGCTCTCC--AACAGCTCC 412
 Db 172 --AsperAlaCysSerSerAlaProGlySerLysProSerSerProAsnHisSer 190

QY 413 CACAGCAGCACTGCTGAATGCTTACTGCTGCTCCCAACATCCCACTGAGATG 472
 Db 191 GlySerAlaSerAlaGluAsnGlyLeuAlaProAlaValProSerThrProAlaGluThr 210

QY 473 CTCCCTGAGACGCGCTCCCTGAGCAGCTCCCAACAGCTTACAGCTTACAGC 532
 Db 211 SerLeuAlaHisArgLeuValAlaArgGlnLysSerAlaAlaProLeuProLeuTyrThr 230

QY 533 TCTCTCTCTGCGCAACATCTCCCTGAGGCTGAGCGCAGCGTCACTGTCACCACTCA 592
 Db 231 SerProSerLeuProAsnLysThrLeuGlyLeuProAlaThr----- 244

QY 593 CACCTACTGCTCCCGGAAAGCTGTGCACA-----CAGCAGAGGCGCGAGGCGAGCC 646
 Db 245 -----GlyProSerAlaGlyThrAlaGlyGlnGlnAspThrGluArgLeuThr 260

QY 647 CTCCAGTCCCTGGCGGAG-----GCTGGCAGCGTACCGGCACTTGTATGATG 691
 Db 261 LeuProAlaLeuGlnGlnAlaArgLeuSerLeuPheProGlyThrHisLeuThrProTyrLeu 280

QY 692 AGCAGCTCTCTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 751
 Db 281 SerThrSer-----ProLeuGlnAlaGlyAspGlyAla 291

QY 752 CAGGCGCATGCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 811
 Db 292 ---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGlnGlnProProAlaGlnAla 310

QY 812 ACCCTACTT-----GCTGGCAGCTTCCAGCGGAGTCCCACTAGTACGGGT 859
 Db 311 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 329

QY 860 GAACGTGTGGCCACAGCATGCGAGCGTAGCGAGAGCTCCCGCGCTGCGCCCTGAGC 919
 Db 330 AspaArgLysSerProSer-----LysHisLysLeuArgGlnHisArgProLeuGly 346

QY 920 CGCACTAGTCTCACCCTGCGCGCAGAGTCCCGAGCGCTGACAGAGTGTGTCATGCA 979
 Db 347 ArgThrGlnSerSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 366

QY 980 CAACAGCACCGACAGTTCCTGAGAGACAGAG-----CAGACAGCTACAG 1027
 Db 367 GlnGlnHisGlnGlnPheLeuGlnLysHisLysGlnGlnPheGlnGlnGlnGln 386

QY 1028 CTGGCAGAGATCTCTACCAACAGACAGGAGCTGCCAGCGCAGCCACACCCCTGAG 1087
 Db 387 MetAsnLysLysLysLeuProLysProSerGluProAlaArgGlnProGlnSerHisProGlu 406

QY 1088 GAGCAGAGAGAGAGCTGACGAGACAGAGAGTCTTCTGGGGGAGGAGCCCTG--- 1144
 Db 407 GlnThrGlnGlnGlnLysLeuArgGlnHisGln---AlaLeuLeuAspGluProTyrLeuAsp 425

QY 1145 ACCATGCCCGGAGGGGCTCCACAGAGGTGAGCAGACAGAGAGCTGAGAGAGAG 1204
 Db 426 ArgLeuProGlyGlnLysGlnAlaHisAlaGlnAlaGlnValGln---ValLysGlnGlu 444

QY 1205 GACGAGAGAGAGATGGGAGGAGAGAGAGATTCACAGTTCAGTTCAGAGAGAGCC 1264
 Db 445 ProIleGlnSerAspGluGlnGlnAlaGlu-----ProProArgGluVal 459

QY 1265 GAGAGTGTGCTGAGAGAGGCGCGGACTTGAGAGAGCTGTGCTGATACAAAACATG 1324
 Db 460 GlnProGlnGlnArgGln---ProSerGlnGlnGlnLeuLeuPheArgGlnGlnAlaLeu 478

QY 1325 TTCTCAGATGCCCAACCGCTGCAACCTTTCGAGGTATACAGAGCGCCCTCAGCTGGCC 1384
 Db 479 LeuLeuGlnGlnGlnAlaGlnLysLysGlnLeuArgAsnTyrGlnAlaSerMetGlnAlaAla 498

QY 1385 ACTGTGCC-----CACCAAGCCCTGGGCGCTTACCAATCTCCCTGCT 1429
 Db 499 GlyLeuProValSerPheGlyHisArgProLeuSerArgAlaGlnSerSerProAla 518

QY 1430 GCC-----CCTGGGCGCATGAGAGACCCCAACCAACCCGTCAGACACTCTTC 1480
 Db 519 SerAlaThrPheProValSerValGlnGluProProThrLysPro-----ArgPhe 535

QY 1481 ACCCAAGTGGTCTGACAGACGCTGATGAGACAGCGATGATGAGGAGGAGGAGACA 1540
 Db 536 ThrThrGlnLeuValIlyTrpAspThrLeuMetLeuLysHisGlnCysThrCysGlySerSer 555

QY 1541 CACGTGACCCCTGAGACATGCTGGCCGAGATCCAGAGCATCTGCTCCGCTGACAGAGACA 1600
 Db 556 SerSerHisProGlnHisAlaGlnAlaGlyAlaGlnSerLysThrArgLeuGlnGluThr 575

QY 1601 GCGCTGCTTACCAAGTGCAGCGGATCCGAGGTGCAAGGCGCATGATGAGATCCAG 1660
 Db 576 GlyLeuArgGlyLysCysGlyCysIleArgGlyArgLysAlaThrLeuGlnGlnGln 595

QY 1661 ACAGTGCAGCTGAAATACACACCTGCTGATGAGACAGCTCCCAACCGGCGAGAG 1720
 Db 596 ThrValHisSerGlnAlaHisThrLeuLeuLysGlyThrAsnProLeuAsnArgGlnLys 615

QY 1721 CTAGACAGCAAGAGTGTCTGCTCCATCAGCCAGAGAGATGTAAGTGTCTGCTGCTGCT 1780
 Db 616 LeuAspSerLysLysLysLeuLeuGlySerLeuAla---SerValPheValArgLeuProCys 634

QY 1781 GGGGCGATCGGGGTGACAGAGTACACCGTGTGATGATGATGATGATGATGATGATG 1840
 Db 635 GlyGlyValGlyAlaAspSerAspThrLysLysValHisSerAlaGlnAlaAla 654

QY 1841 CGCATGGCAGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1900
 Db 655 ArgLeuAlaValGlyCysValValGlnLeuValPheLysValAlaThrGlnGlnLys 674

QY 1901 AATGATTTGCATCATCCGCGCCCGAGACACACCGCAGAGATCCACAGCATGAGGA 1960
 Db 675 AsnGlyPheAlaValAlaArgProProGlnHisHisAlaGlnGlnSerThrPrometCyl 694

QY 1961 TTCTGCTTCTTCACTGTGAGCCATACCGCAAACTCTTACAGCAGAGAGTGAACGTG 2020
 Db 695 PheCysTyrPheAsnSerValAlaValAlaAlaLysLeuLeuGlnGlnArgLeuSerVal 714

```

QY 2021 GGCAGGCTCTATCTGGACACTGGACATTCACCATGGCAATGGCAGCCAGCGCTTC 2080
    |||:::|||||
Db 715 SerLysIleLeuIleValIleSptrIpaSerValHisHisGlyAsnIleYthrGlnGlnAlaIpaHe 734
QY 2081 TACAAATGACCCCTCTGGCTCTACATCTCTGATCGGATGACAGCGGAATCTCTT 2140
    |||:::|||||
Db 735 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspSpGlyAsnIpaHe 754
QY 2141 CCAAGCTCTGGGGCTCTCCAGAGAGTGGTGGAGAGCAAGCGCTGGGATACATGTGAAC 2200
    |||:::|||||
Db 755 ProGlySerGlyAlaIpaProSpGlyValGlyThrGlyProGlyValGlyIleValAsnValaSn 774
QY 2201 GTCGACATGGACAGAGAGTGTGGACCCCCCATTTGGAGAGCTGGAGTACCTTACAGCTTC 2260
    |||:::|||||
Db 775 MetAlaPheThrGlyIleGlyLeuAspProPheMetGlyAspAlaIleTyrIleValAlaIpaHe 794
QY 2261 AGGACAGTGGTGTATGCCCTTGGCCAGAGTTCTCACCTGATGTGGTCTTACGTCTCGCCG 2320
    |||:::|||||
Db 795 ArgThrValValMetProIleAlaSerGluPheAlaProAspValValLeuAlaSerSer 814
QY 2321 GGGTTTGATGCTGTGTAAGACATCTCTCTCTGGTGGCTACTCTGTACCCGCCAGA 2380
    |||:::|||||
Db 815 GlyPheAspAlaValGlyIleGlyHisProThrProLeuGlyIleTyrAsnLeuSerAlaIpaG 834
QY 2381 TGTGTTGGGACCTTGAACAGAGCTGATGACCTGGAGACCCCTGGAGGGGGGGGCTGGCTGCC 2440
    |||:::|||||
Db 835 CysPheGlyTyrIleuThrLysGlnLeuMetGlyLeuAlaGlyIleYthrGlyIleValLeuAla 854
QY 2441 CTGAGAGGAGGAGCCATGACTTGAACCCCATCTGATGATCCCTGGAACCTGTGTCTGGCT 2500
    |||:::|||||
Db 855 LeuGlnGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGlnAlaGlyValSerAla 874
QY 2501 CTGCTCAGTGTAAAGCTGCAAGCCCTTGATGAGGACGCTTTCAGCAAAAGCCCAACATC 2560
    |||:::|||||
Db 875 LeuLeuGlyAsnGlyIleuLeuAspProLeuProGlyIleValLeuGlnGlnArgProAlaIpa 894
QY 2561 AACGGCGCCAGCCATAGCAAGATCATCGAGATCCAGAGCAAAAGCAGAGCTGTGG 2620
    |||:::|||||
Db 895 AsnAlaValArgSerMetGluIleuValMetGlyIleHisSerLysTyrTrpArgCysIleu 914
QY 2621 CAGAAATTCGCCCTGCTGGGCGCGGCTCCCTGCGAGGGGCCCAAGCAGATGAGACCGGAA 2680
    |||:::|||||
Db 915 GlnArgThrThrSerThrAlaGlyArgSerLeuIleGlnAlaGlnThrCysGlnAsnGlu 934
QY 2681 GAAGCCGAAT-GTGAACGCCATGGCCCTTGTCTGGTGGGGGGCGCAACAGGCCCAAGCT 2739
    |||:::|||||
Db 935 GlnAlaGlnThrValThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu--- 953
QY 2740 GCGGACCCCGGGACACAGCCCGGCGGAGAGAGAGCCATGAGAGAGCTGCC 2799
    |||:::|||||
Db 954 -----LysArgProAspGlnGluProMetGlnGlnProPro 966
QY 2800 CTG 2802
    |||
Db 967 Leu 967

```

```

; SEQ ID NO 7
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Human
US-09-817-538-7

Alignment Scores:
Pred. No.: 1.6e-124 Length: 967
Score: 2685.00 Matches: 567
Percent Similarity: 71.28% Conservative: 118
Best Local Similarity: 59.00% Mismatches: 196
Query Match: 50.22% Gaps: 81
DB: 10 Gaps: 23

US-09-502-945-2 (1-2885) x US-09-817-538-7 (1-967)
QY 2 GAATTCCTGTGTCGAAAGTCAAGAGCCACACAGCGGGGCTCAACATTCCTCCCA 61
    |||:::|||||
Db 60 GluPheValLeuAsnLysLysAlaLeuAlaHisProAsnLeuHisCysIleSer 79
QY 62 CAGCACCCCAATGCTGG-----GGAGCCACATGCTTCTTGGACAGAGTTCCCT 115
    |||:::|||||
Db 80 SerCysProArgTyrTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerPro 99
QY 116 CCCAGAGCGGGCCCCCTGGAGCCCTCCCTCTACAACTGCTTGGCGCCCTAC 175
    |||:::|||||
Db 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuIleMetTyr 116
QY 176 GACAGTCAGAGAGACTTCCCTCCGCAAAACAGCCTGTGAACCCCAACTTGAAGTGCAT 235
    |||:::|||||
Db 117 AspAlaIleAspAspPheProLeuArgTyrThrAlaSerGlnProAsnLeuLysLeuArg 136
QY 236 TCAAGGCTAAACAGAAAGTGGCTGAGCGAGAGAGCATCCCTCGTGGCAAGAT 295
    |||:::|||||
Db 137 SerArgLeuLysGlnLysValAlaGlnArgArgSerSerProLeuLeuArgLysAsp 156
QY 296 GGGACTGTATATGACACCTTAAAGAGAGAGCTGTGAATACAGAGTCCCGGCTGGG 355
    |||:::|||||
Db 157 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 171
QY 356 GCGTGCCTCGGTGTAAACAGCGCACCCGGCTCGGCGCCAGCTCTCC--AACAGCTCC 412
    |||:::|||||
Db 172 ---AspSerAlaCysSerSerAlaProGlySerGlyProSerSerProAsnAsnSerSer 190
QY 413 CACAGCACCATGCTGAGAATGCTTACTGTGCTAGTCCAGTCCCAACATGCCCACTGAGATG 472
    |||:::|||||
Db 191 GlySerValSerAlaGlnAsnGlyIleAlaProAlaValProSerLleProAlaGluThr 210
QY 473 CTCCCTCAGACACCGACCCCTCCCTGGACAGACTGCCCAACACAGTTACAGCTTACAGG 532
    |||:::|||||
Db 211 SerLeuAlaHisArgLeuValAlaArgGlnGlySerAlaAlaProLeuProLeuTyrThr 230
QY 533 TCTCTCTCTGCGCCAAACATCTCCCTAGGGCTGACGGCCAGCGTCACTGCAACATCA 592
    |||:::|||||
Db 231 SerProSerLeuProAsnLleThrLeuGlnGlyLeuProAlaThr----- 244
QY 593 CACCTCAGTGCCTCCCGAAGCTGTGCACA-----CAGCAGAGAGCGGAGAGCGCC 646
    |||:::|||||
Db 245 -----GlyProSerAlaGlyThrAlaGlnGlnAlaSprThrClnArgLeuThr 260
QY 647 CTCAGATCCCTCGGCGAG-----GGTGGACAGCTACCGGCAAGTTTCATG 691
    |||:::|||||
Db 261 LeuProAlaLeuGlnGlnArgLeuSerLeuPheProGlyThrHisLeuThrProTyrLeu 280
QY 692 AGCAGATCTCTATTCCTGGCTGCTGCTGGGCGTGGGCACTGGAGGGAGCGGAGCCCC 751
    |||:::|||||
Db 281 SerThrSer-----ProLeuGlnArgAspGlyAla 291
QY 752 CACGGGAGATGCCCTCTGTCAGAGCATGTGCTGTGAGAGAGCGGCGGAGCAGAGC 811
    |||:::|||||
Db 292 ---AlaHisSerProLeuLeuGlnHisMetValLeuLeuGlnGlnProProAlaGlnAla 310
QY 812 ACCCTCATTT-----GCTGTGCCACTCCAGGCGGCACTGCCCACTAGTACGGGT 859

```



```
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-414US1
; CURRENT APPLICATION NUMBER: US/10/072.094
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/296,296
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 87
; LENGTH: 1069
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-094-87

Alignment Scores:
Pred. No.: 1.09e-123 Length: 1069
Score: 2668.00 Matches: 570
Percent Similarity: 70.88% Conservative: 114
Best Local Similarity: 59.07% Mismatches: 203
Query Match: 49.91% Indels: 79
DB: 9 Gaps: 21

US-09-502-945-2 (1-2885) x US-10-072-094-87 (1-1069)
QY 2 GAATTCCTCTGTGAGTCA-----AAGAGCCACACAGCGCGCTCAACCATTC 55
   |||||
Db 152 GtupheuleuSerLysSerLalaThrLysAspThrProthraSnglyLysAsnHisSer 171
   ::::
QY 56 CTCACAGACACCCAAATGCTG-----GGAGCCACCATGCTTCTTGAGCAGAGT 109
   ::::
Db 172 ValSerArgHisProLysLeuTryptThrAlaHisHisThrSerLeuAspLysSer 191
   ::::
QY 110 TCCCTCCCAAGAGCGCGCCCTGGAGCGCTCCCTCTCAAACTGCTTGGCTGG 169
   |||||
Db 192 SerProLeuSer-----GlyThrSerProSerTyrLysTyrThrLeuProGly 208
   |||||
QY 170 CCTTCAGACAGTGCAGACACTCCCTCCGCAAAACAGCTTGAGCCCAACTTGA 229
   |||||
Db 209 AlaGlnAspAlaLysAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLys 228
   |||||
QY 230 GTGCGTTCAAGGCTAAACAGAGGTGCTGAGCGAGACAGTCCCTCTGCGTGC 289
   |||||
Db 229 ValArgSerArgLeuLysGlnLysValAlaGluArgArgSerProLeuLeuArgTyr 248
   |||||
QY 290 AAGATGGGACTGTTATAGACCTTTAAGAAGAGCTGTGAGATACAGAGTCCCGG 349
   |||||
Db 249 LysAspGlyAsnValValThrSerPheLysArgMetPheGluValThr----- 265
   |||||
QY 350 CCTGGGCGCTGCTGTAACAGCGACCGCGCTCGGCGCCAGCTTCCCAAGC 409
   |||||
Db 266 -----GluSerSerValSerSerSerProGlySerGlyProSerSerProAsn--- 282
   |||||
QY 410 TCCACAGACACCATGCTGAGAAATGGCTTACTGGCTTACGCCCAACATCCCACTAG 469
   |||||
Db 283 -----AsnGlyProThrGlySerValThrGluAsnGlnThrSer 295
   |||||
QY 470 ATGCTCCCT-----CAGCAGCGAGCCCTCCCTCTG 499
   |||||
Db 296 ValLeuProThrProHisAlaGluGlnMetValSerGlnGlnArgLysLeuLeuHis 315
   |||||
QY 500 GACAGCTCCCCACCACTTACAGCTTACAGCTCTCTTGGCCCAAGATGCCCTA 559
   |||||
Db 316 GlnAspSerMetAsnLeuLeuSerLeuTyrThrSerProSerLeuProAsnLethrLeu 335
   |||||
QY 560 GGGCTGCAGCGCAGGTCATGTCACCAACTACACACTACAGCTCCCGCAAGCTGTGC 619
   |||||
Db 336 GlyLeuProAla-----ValProSerGlnLeuAsnAlaSerAsnSerLeuLys 351
   |||||
QY 620 ACACAGCAGGAGGCCGAGAGCGAGCGCTCCAGTCCCTGCGGAGGTGGACGCTGACC 679
   |||||
Db 352 GluLysGlnLysCysGlu-----ThrGlnThrLeuArgGlnGlyValProLeuPro 368
```

```
QY 680 GGCAGATTC-----ATGAGCAATCTATTCTGCTGCTGCTGCTGCGGCTG 727
   |||||
Db 369 GlyGlnTyrGlnLysSerLethProAlaSerSerSerHisPro-----HisVal 384
   |||||
QY 728 GCATGGAGGGCGAGGGAGGCCCGCCAGCGATGCTCCCTGCTGCAAGATGCTGTG 787
   |||||
Db 385 ThrLeuGlnLysProProAsnSerSerHisGlnAlaLeuGlnHisLeuLeu 404
   |||||
QY 788 CTGAGAGCGCCCGCAGACAGACAGCCCTCATTCCT-----GTGCCATCTCCAGGCGAG 841
   |||||
Db 405 LysGlnGlnMetArgGlnGlnLysLeuLeuValAlaGlyValProLeuHisProGln 424
   |||||
QY 842 TCCCCACTAGTACGGGTGAAGCTGTGTGACCCAGCATGCGGAGCGAGTACGCACTCCG 901
   |||||
Db 425 SerProLeuAlaThrLysGlnArgGlySerProGlyLethArgLysThrHisLysLeuPro 444
   |||||
QY 902 CGGATCGGCGCTGAGCGCCGACATGCTTCATCCAGCGCTGCGGAGATGCCAGGCGCTG 961
   |||||
Db 445 ArgHisArgProLeuAsnArgThrGlnSerAlaProLeuProGlnSer-----ThrLeu 462
   |||||
QY 962 CAGCAGCTGTCATGCAACACAGACAGCAGCATTCCTGGAGAGCAGAG-----CAG 1015
   |||||
Db 463 AlaGlnLeuValIleGlnGlnGlnHisGlnLysGlnLysGlnLysGlnTyrGln 482
   |||||
QY 1016 CAGCAGCTACAGCTGGGAGATCTCACAAGACAGGAGGAGCTGCGCCAGCGCCAGCC 1075
   |||||
Db 483 GlnGlnIleHisMetAsnLysLeuLeuSerLysSerIleGlnGlnLysGlnProGly 502
   |||||
QY 1076 ACCCACTTGAAGACAGACAGAGAGAGCTGACGAGCAGCAGAGAGCTTCTGCGGAG 1135
   |||||
Db 503 SerHisLeuGlnGlnAlaGlnGlnGlnLeuGln-----GlyAsp 515
   |||||
QY 1136 GGAGCGCTGACCATGCGCGGAGGCGTCCACAGAGATGAGACACAGAGAGACTG 1195
   |||||
Db 516 GlnAlaMetGlnGlnAspArgAlaProSerSer---GlyAsnSerThrArgSerAspSer 534
   |||||
QY 1196 GAGAGGAGGAGCAGAGAGAGAGATGGGAGGAGAGAGATTCATCCAGGTTAAGAC 1255
   |||||
Db 535 SerAlaCysValAspAspThrLeuGly-----GlnValGlyAlaValLysValLysGln 552
   |||||
QY 1256 GAGAGGGCGCAGAGTGGTCTGAGAGAGGCGCCCATTTGAGAGGCGCTGCTGATAC 1315
   |||||
Db 553 GluProValAspSer-----AspGlnAspAlaGlnIleGlnIleMetGlnSerGlyGlu 570
   |||||
QY 1316 AAAAAGCTGTCAGATGCCCAACCGCTGCACACT-----TTGCAGGTG 1360
   |||||
Db 571 GlnAlaAlaPheMetGlnGlnProPheLeuGlnProThrHisThrArgAlaLeuSerVal 590
   |||||
QY 1361 TACCAGCGCCCTTCAGCTGCGCAGTGTG-----CCCAACCAAGCCCTGCGG 1408
   |||||
Db 591 ArgGlnAlaProLeuAlaAlaValGlyMetAspGlyLeuGlnLysHisArgLeuValSer 610
   |||||
QY 1409 CGTACCCCAATCTCCCTGCTGCGCCCTGGGGGCGATGAAGAAGCCCAAGACCAACCGCTC 1468
   |||||
Db 611 ArgThrHisSerSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeu 630
   |||||
QY 1469 AAGCACTCTTCACCAACAGTGTGCTACAGACAGCTTCATGCTAAAGCAACAGTGCATG 1528
   |||||
Db 631 GlnProGlySerAlaThrGlyIleAlaLysTyrAspProLeuMetLeuLysHisGlnCysVal 650
   |||||
QY 1529 TGCAGAAACACACAGCTGACATGCTGAGCCGCGATCCAGACATCTGCTCCGCG 1588
   |||||
Db 651 CysGlnLysAsnSerThrThrHisProGlnHisAlaGlyLysArgIleGlnSerIleTyrSerArg 670
   |||||
QY 1589 CTGCAGAGACAGGCTGCTTAGCAAGTGCAGCGAGTCCAGAGTGCCTCAAAAGCCAGCTA 1648
   |||||
Db 671 LeuGlnGlnThrGlyLeuLeuAsnLysCysGlnLysArgIleGlnGlnArgLysAlaSerLeu 690
   |||||
QY 1649 GATAGATCCAGAGAGTGCATGATACACACCTGCTCTTGGAGACAGTCCCTC 1708
   |||||
Db 691 GlnGlnIleGlnLeuValHisSerGlnHisHisSerLeuLeuTyrGlnThrAsnProLeu 710
```

```
QY 1709 AACGGCAGAGCTAGACAGAGATTGCTGCTCCATCCAGACGAGATGATGCT 1768
   ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 711 AspGlyGlnIleuAspProArgIleLeuLeuGlyAspAspSerGlnIleuPhePheSer 730
QY 1769 GTGCTGCTGTGGGGGCGGGGTGGACAGTACACCGCTGTGAATGATGACATCC 1828
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 731 SerLeuProCysGlyGlyLeuGlyValAspSerAspThrIleTrpAsnGlnIleuHisSer 750
QY 1829 TCCAGTCTGTGGCAGAGTGGGCTGCTGCTGAGCTGGGCTTCAAGTGGCTGCA 1888
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 751 SerGlyAlaIleAlaIleAlaValGlyCysValIleLeuIleuAlaSerValAlaSer 770
QY 1889 GGAGAGCTCAAGATGATTTGGCATATCCGGCCCCAGACAGCAGCCGAGATCC 1948
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 771 GlyGlnLeuLysAsnGlyPheAlaValAlaArgProProGlyIleHisAlaGlnGlyUser 790
QY 1949 ACAGCCATGGGATTCGCTTCTTCAACTGTGTGCCATCACCGCAAACTCCTACAGCAG 2008
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 791 ThrAlaMetGlyPheCysPhePheAsnSerValAlaIleThrAlaLysTyrLeuArgAsp 810
QY 2009 AAGTTGACGCTGGGAGGCTCATGCTGAGCTGGAGCTGGACATTCACCATGGCATGGCACC 2068
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 811 GlnIleuAsnIleSerIleLeuIleValAlaAspLeuAspAlaHisIleGlyAsnGlyThr 830
QY 2069 CAGCAGCGCTTCTACATGACACCCCTGTGCTTACATCTCTGTGCATCGCTATGACAAAC 2128
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 831 GlnGlnAlaPheTyrAlaAspProSerIleLeuTyrIleSerIleuHisArgTyrAspGln 850
QY 2129 GGAACCTCTTTCCAGGCTCTGGGGCTCCTGAAGAGTGTGTGGAGACAGCCGCTGGGG 2188
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 851 GlyAsnPhePheProGlySerGlyAlaProAsnGlnValGlyIleGlyLeuGlyGlnGly 870
QY 2189 TACATGTGAACGTGGCATGGACAGAGGTGTGGACCCCGCCATTTGAGAGCTGGAGTAC 2248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 871 TyrAsnIleAsnIleAlaTrpThrGlyGlyLeuAspProProMetGlyAspValGlyTyr 890
QY 2249 CTTCACGCTTCAAGCAGTGTGATGCCATTCGCCACAGATTCCTACCTGATGTGCTC 2308
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 891 LeuGlnAlaPheArgThrIleValLysProValAlaLysLeuPheAspProAspMetVal 910
QY 2309 CTAGTCTCGCGGGTGTATGCTGTGAAGACATCTGCTCTGCTGGTGGCTACTCT 2368
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 911 LeuValSerAlaGlyPheAspAlaLeuGlnGlyHisThrProProLeuGlyGlyTyrLys 930
QY 2369 GTGACCGCAGATGTTTGGCCACTTGACACGAGCTGTGATGACCCCTGGAGGGGGCGG 2428
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 931 ValThrAlaLysCysPheGlyHisIleLeuThrLysGlnLeuMetThrLeuAlaAspGlyArg 950
QY 2429 GTGGTCTGTGGCTTGAGGAGGCGCATGACTTGACCGCATCTGTATGCTGTGAAGCT 2488
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 951 ValValIleuAlaLeuGlnGlyGlyHisAspLeuThrAlaIleCysAspAlaSerGlnAla 970
QY 2489 TGTGTTCTGGCTGTGCTAGTGAAGCTGACACCTTGATGATGAGCAGTCTTGACGAA 2548
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 971 CysValAsnAlaLeuLeuGlnLysAsnGlnLeuGlnProLeuAlaGlnIleuHisGln 990
QY 2549 AAGCCCATCATCAACGACAGTGGCCAGCTGTAGAGAAATGTCAGATCCAGACCAACAC 2608
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 991 SerProAsnMetAsnAlaValIleSerLeuGlnLysIleIleIleIleGlnSerLysTyr 1010
QY 2609 TGGAGCTGTGTGCAAGATTCCGCTGTGCTGGCGCGGTCTGCTGGAGGGGCCCAAGCA 2668
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1011 TrpLysSerValArgMetValAlaValAlaProArgGlyCysAlaLeuAlaGlyAlaGlnLeu 1030
QY 2669 GGTGAGACCGAAGACCGCAAT-GTGAACGCCATGCGCTTGTGTTGGTGGGGCCGAA 2727
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1031 -----GlnGlnGlnIleuThrValSerAlaLeuAlaSerLeuThrValAspValGln 1048
QY 2728 CAGGCCCAAGCTGCGGACGCGCGGACACACAGCCCGGACAGAGAGCCCATGAG 2787
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1049 GlnProPheAla-----GlnGlnAspSerArgThrAlaGlyIleuProMetGln 1064
QY 2788 CAGGAGCTTGCCCTG 2802
```

```
Db 1065 GlnGlnProAlaLeu 1069
RESULT 9
US-10-072-094-89
; Sequence 89, Application US/10072094
; Publication NO. US20030025538A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, DONALD
; APPLICANT: LORENZI, MATTHEW
; APPLICANT: ATTAR, RICARDO
; APPLICANT: GOTTARDIS, MARCO
; TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
; FILE REFERENCE: 3053-4145US1
; CURRENT APPLICATION NUMBER: US/10/072,094
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/298,296
; PRIOR FILING DATE: 2001-06-14
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 89
; LENGTH: 1011
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-072-094-89

Alignment Scores:
Pred. No.: 2,51e-119 Length: 1011
Score: 2579.00 Matches: 540
Percent Similarity: 72.05% Conservative: 107
Best Local Similarity: 60.13% Mismatches: 179
Query Match: 48.24% Indels: 72
DB: Gaps: 19

US-09-502-945-2 (1-2885) x US-10-072-094-89 (1-1011)
QY 2 GAATTCCTTGTGCAAGTCA-----AAGGAGCCCAACACAGCGGCTCAACCATTC 55
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 GlnPheLeuLeuSerIleLysSerAlaIleThrLysAspThrProThrAsnGlyLysAsnHisSer 168
QY 56 CTCGCCACAGCAGCCCAAAATGCTGG-----GAGGCCCATGCTTCTTTGGACAGAGT 109
   ::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 169 ValSerArgHisProLysLeuTyrThrAlaAlaHisIleThrSerLeuAspGlnSer 188
QY 110 TCCCTCCCCAGAGGCGGCCCGCTGGGAGCGGCTCCCTGACAAACGTCCTTGGCTGGG 169
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 189 SerProLeuSer-----GlyThrSerProSerIleLysTyrThrLeuProGly 205
QY 170 CCTACAGACAGTGAAGCAGACTCCCTCCGCAAAACAGCCTGAACCCAACTTGAA 229
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 206 AlaGlnAlaPheAlaLysAspAspPheProLeuArgLysThrAlaSerGlnProAsnLeuLys 225
QY 230 GTGCGTTCAAGGCTAAACAGAGTGTGCTGAGCGAGAGAGCAGTCCCTCTGCTGCTGC 289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 226 ValArgSerArgLeuLysGlnLysValAlaGlnArgArgSerSerProLeuLeuArgArg 245
QY 290 AAGGATGGGACTTTATTTAGCACCCTTAAGAGAGAGGCTGTGAGTTCACAGGTGCCGG 349
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 246 LysAspGlyAsnAlaValIleThrSerPheLysLysArgMetPheGlnValIleThr----- 262
QY 350 CTGGGGGCTGCTGCTGTGTACACAGCAGACCGCGCTCCGGCCGAGCTCTCCCAACAGC 409
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 -----GlnSerSerValIleSerSerSerProGlySerGlyProSerSerProAsn---- 279
QY 410 TCCACAGCAGCATCGCTGAGATGAGCTTACTAGCTCAGTGTCCCAACATCCCACTGAG 469
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 -----AsnGlyProThrGlnGlySerValIleThrGlnLysGlnIleuThrSer 292
QY 470 ATGCTCCCT-----CAGCAGCGAGCCCTCCTCTG 499
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 293 ValIleuProThrProThrAlaGlnGlnMetValSerGlnArgIleuLeuLeuHis 312
QY 500 GACAGCTCCCCCAACAGATTCAACGCTCTTAACGCTCTTCCTGCGCAACATCTCCCTA 559
```

D	b	313	gLuAspSerMetAsnLeuLeuSerLeuTyrThiSerProSerLeuProAsnIleLeuHis	332
O	y	560	GGGCTGACGAGCCACGGTACTGTCCAACTCAACCTCACTGCTCCCGAAGCTTGC	619
D	b	333	gLyLeuProAla-----ValProSerGlnLeuAsnAlaSerAsnSerLeuLys	348
O	y	620	ACACGAGGAGGAGGCGGAGGAGCAGGCGCTCCATGCTCCGGGAGGGGTGGACGGTAC	679
D	b	349	gLuLysGlnLysCysGln-----ThiGlnIleuArgGlnIleValProLeuPro	365
O	y	680	GGCAAGTTC-----ATGAGCACAATCTCATTCCTGGCTGGCTGCTGGCGGTG	727
D	b	366	gLyGlnTyrGlnLysLeuSerIleProAlaSerSerHisPro-----HisVal	381
O	y	728	GCATGGAGAGGCGACGGGAGGCGCCCGACGGGATGCTCCCTGTGAGCATGTGCTTG	787
D	b	382	ThiLeuGlnLysProProAsnSerSerHisGlnAlaLeuLeuIleHisLeuLeu	401
O	y	788	CTGAGCAGGCGCCGAGCAGAGCAGACCTCATTTGCT-----GTGCACTCCACGGGAG	841
D	b	402	LysGlnGlnMetArgGlnGlnLysLeuLeuValAlaGlyLysValProIleHisProGln	421
O	y	842	TCCCACTAGTACAGGCGGTGAACGTGTGGCCACACAGATCGGAGCGGTAGCACTCCG	901
D	b	422	SerProLeuAlaThrLysGlnTyrGlnIleSerProGlyIleArgIleThiHisLysLeuPro	441
O	y	902	CGGCACTGGCCCTGAGCGCGGACACTTCCTCAACCGCTGGCGGAGTCCCGACGGCCGTG	961
D	b	442	ArgHisArgProLeuAsnAlaArgThrGlnSerAlaProLeuProGlnSer-----ThiLeu	459
O	y	962	CAGCAGCTGTGATGCAACAACACACAGCAGCACTTCTGTGAGAAAGCAGAG-----CAG	1013
D	b	460	AlaGlnLeuValIleGlnGlnGlnHisGlnIlePheLeuGlnLysGlnLysGlnTyrGln	479
O	y	1016	CAGAGCTACACAGCTGGGCGAGATGCTCCACAAAGGAGAGCTGGCCAGGAGGCCAAC	1075
D	b	480	GlnGlnIleHisMetAsnLysLeuLeuSerLysSerIleGlnIleLysGlnProGly	499
O	y	1076	ACCCACCTGTGAGGAGACAGAGAGAGACTACGAGCAGACAGAGAGGTCTTGTGGGGAG	1135
D	b	500	SerHisLeuGlnGlnIleAlaGlnLysGlnLeuGln-----GlyAsp	512
O	y	1136	GGAGCCCTGACCAATGCCCGGGAGGGCTCCACAGAGATGAGAGCACACAGAAACCTG	1195
D	b	513	GlnAlaMetGlnLysAspArgAlaProSerSer--GlyAsnSerThrArgSerAspSer	531
O	y	1196	GAGGAGGAGGAGGAGAAAGATGGGGAGGAGAGAGAGAGATGTGCATCCAGGTAAAGAC	1255
D	b	532	SerAlaCysValaAspArgThrIleuGly-----GlnValGlyAlaValLysValLysGln	549
O	y	1256	GAGCAGGGCGAGAGTGTGCTGTAGAGAGGGGCGCCGACTGTGGAGAGGCTGTGCTGATAC	1315
D	b	550	GluProValaAspSer-----AspGluAspAlaGlnIleGlnIleMetLysSerLysGln	567
O	y	1316	AAAAAATGTGTCTGAGATGCCCAACCGCTGCACACT-----TTGGAGGG	1366
D	b	568	GlnAlaAlaPheMetGlnGlnProPheLeuGlnIleProThrHisThrArgAlaLeuSerVal	587
O	y	1361	TACCAAGCGCCCTGACGCTGGCCACTGTG-----CCCAACCAAGCCCTGGGC	1408
D	b	588	ArgGlnAlaProLeuAlaAlaValAlaGlyMetAspGlyLeuGlnLysHisArgLeuValSer	607
O	y	1409	CGTACCCAAATGCTCCCTGCTGCTGCCCTGGGGGCGATGAAGAACCCCGACACCAACCTGC	1468
D	b	608	ArgThrHisSerProAlaAlaSerValLeuProHisProAlaMetAspArgProLeu	627
O	y	1469	AAGACACCTTCAACCAAGATGTGTGTACAGACCTTGATGTAAAGACACAGTGCATG	1528
D	b	628	GlnProGlnSerAlaThrGlnGlyIleAlaTyrAspProLeuMetLeuLysHisGlnLysVal	647
O	y	1529	TGGGGAAACACACAGCTGACACCTGAGCATGTGGCGGATGCCAGAGCATCTGTGTCGGG	1588

D	b	648	CysGlsYnsnserThrTlnHnSproGlnHnIsaGlyValGlnleolnserLletrpserArg	667
Q	y	1589	CTCGACGAGACAGGCCCTGTTACGACGTGGACGGATCCGAGTGGCAACGACGCTA	1648
D	b	668	LeuGlnGlnTlrrGlyLeuLeuAsnLysCysGlnArgGlnleolnGlyArgLysAsaSerLeu	687
Q	y	1649	GATGAGATCCAGACAGGACGCTCGAATACCAACGCTCTGATAGGACGACGTGCTCCCTC	1708
D	b	688	GlnGlnIleGlnLeuValHnSserGlnHnIsnSerLeuLeuTyrGlnLysHnSproLeu	707
Q	y	1709	AACGGGAGAGCTAGACAGCAGAAAGTTGCTGGTCCATCAGCCAGAGATGTATGCT	1768
D	b	708	AspGlnGlnLysLeuAsnProArgGlnLeuLeuGlnLysAspSercLlnLysSpheser	727
Q	y	1769	GTGCTGCTTGTGGGGGACCTGGGCTGGAGACAGACGCTGGAGATAGATGTGACCTC	1828
D	b	728	SerLeuProCysGlnGlyLeuGlnLysValAspSerAspHnLletrpSmGlnleuHnSer	747
Q	y	1829	TCCAGCTGTGTGCGCATGGCAGATGGGTGGCTGCTGGAGCTGGGCTTCAAGGTGGCTCA	1888
D	b	748	SerGlnLlnAlaAlaArgMetAlaValGlyCysValIleGlnleuHnAserLysValAlaSer	767
Q	y	1889	GGAAGGCTCAAGATGGATTTGGCATATCCGGGCCCCAGACACCCGCGAGAAATCC	1948
D	b	768	GlyGlnLeuLysAsnGlnPheAlaValAlaArgProGlnLynHnIsnIsaGlnleuHnSer	787
Q	y	1949	ACAGCATGGGATTTCTGCTCTTCAATCTGTAGCCATACCGCAAAATCCTACAGCAG	2008
D	b	788	ThrAlaMetGlyPheCysPhePheAsnSerValAlaIleThrAlaLysTyrLeuHnArgS	807
Q	y	2009	AAGTTGAACGTGGGACAGGCTCTCATCTGTGGATGGGACATTCACCATGGACATGGCAC	2068
D	b	808	GlnLeuAsnLleSerLysIleLeuLleValAspLeuAsnValHnHnIsGlnLysnGlnLyr	827
Q	y	2069	CAGCAGGGGCTCTACATGTACCCCTGTGCTCTTAATCTGTGATGTGATGATGATGACAC	2128
D	b	828	GlnGlnAlaPheTyrAlaAspProSerLleLeuTyrIleSerLeuHnIsArgTyrSpcL	847
Q	y	2129	GGGAATCTTTTCCAGGCTCTGGGGCTCTGAAAGAGTTGGTGGAGACCGCTGGGG	2188
D	b	848	GlyAsnPhePheProGlnSerGlyAlaProAsnGlnValGlnTyrGlnLysGlnLysGln	867
Q	y	2189	TACACATGACGTCGGATGGACAGAGAGTGTGGACCCGCCCATTGGAGACGTGGAGTAC	2248
D	b	868	TyrAsnLleAsnLleAlaTrpHnGlyGlyLeuAsnProTomeCylAspValGlnTlrr	887
Q	y	2249	CTTACAGGCTTCAGACAGACAGTGTGATGTGCCATTTGCCACGAGTTCTTCACTGATGTGCT	2308
D	b	888	LeuGlnAlaPheArgHnLleValLysProValAlaLysGlnPheAspSercMetVal	907
Q	y	2309	CTAGTCTCCGCGGGTTATGCTGTGTGAAGACATCTGTCTCTCTGGGTGGCTACTCT	2368
D	b	908	LeuValSerAlaGlnPheAspAlaLeuGlnGlnLysHnSthrProProLeuGlnLysTyrLys	927
Q	y	2369	GTCACCGGACAGATGTTTGGSCACTTGACACGACACACTATGACCTGGCAGGGGCGG	2428
D	b	928	ValHnAlaLysCysPheGlnLysLleThrLysGlnLeuMetThrLeuHnAspLysArg	947
Q	y	2429	GTGGTGTGGCTGGGAGGAGGACGCTGACTTACGCGCCATGTGTGATGCTCTTAAGCT	2488
D	b	948	ValValLeuAlaLeuGlnGlnGlyGlnHnAspLeuThrAlaIleCysAspAlaSerGlnAla	967
Q	y	2489	TGTATCTCGGCTGTGCTCACTGTAAAGCTGCACGCTTGATGAGGACGACTGTGACGCA	2548
D	b	968	CysValAsnAlaLeuLeuGlnLysnGlnLeuGlnProLeuAlaGlnLysPrlleuHnIsGln	987
Q	y	2549	AAGCGCAACATCAACGACGTGGGCGGCACTAGAGAAAGTTCATGAGATCCAGACG	2602
D	b	988	SerProAsnMetAsnAlaValIleSerLeuGlnLysLleIleGlnIleGlnSer	1005
RESULT 10				
US-10-072-094 - 93				
: Sequence 93, Application US/10072094				

[illegible][illegible]

QY 2108 TCTTCATCGCTATGACACGGGAATCTTTCAGGCTGGGGCTCTGAAAGGTT 2167
|||||
DB 555 SerLeuHISArgTYrAspLeuLysnPhenProGlySerGlyAlaProasnGluVal 574
QY 2168 GGTGGAGGACGAGCGGTGGGTACATGTGAAGCGGCGATGGAGAGGCTGGACCC 2227
|||||
DB 575 GlyThrGlyLeuGlyValGlyLeuLysnIleAsnIleAlaTrpThrGlyGlyLeuAspPro 594
QY 2228 CCCATTTGAGACGTGGAGTACCTTACAGGCTTACAGACAGTGTGATGCCATTCGCCAC 2287
|||||
DB 595 PrometGlyAspValGlyLeuGlyLeuAlaPheArgThrIleValLysProValAlaLys 614
QY 2288 GAGTCTCAGCTGATGTGTCTTACTCTCCCGGGTTTGATGCTGTGAAGACATCTG 2347
|||||
DB 615 GluPheAspProAspMetValLeuValSerAlaGlyPheAspAlaLeuGlyLeuIlyThr 634
QY 2348 TCTCTCTGGTGGTGTACCTGTACCGGCAATGTTTGGCCACTTGACCGGCACTG 2407
|||||
DB 635 PropioLeuGlyGlyTyrLysValThrAlaLysCysPheGlyHisLeuThrLysGlnLeu 654
QY 2408 ATGACCTTGGAGGGGGGGGGTGGTGGCTGGCCCTGAGGGAGGCCATGACTTGACGCC 2467
|||||
DB 655 MetThrLeuAlaAspGlyArgValValLeuAlaLeuGlyGlyHisAspLeuThrAla 674
QY 2468 ATCTGTGATGCTCTGAAAGCTGTGTCTGGCTGCTGCTCAGTGTAAAGCTGACGCCCTTG 2527
|||||
DB 675 IleCysAspAlaSerGlyAlaCysValAsnAlaLeuGlyAsnGlyLeuGlyLeuProLeu 694
QY 2528 GATGAGGACGCTTGTACCAAAAGCCCAACATCAACGACGATGGCCGCTAGAGAAAGTC 2587
|||||
DB 695 AlaGluAspIleLeuHisGlnSerProAsnMetAsnAlaValIleSerLeuGlnLysIle 714
QY 2588 ATCGAGATCCAGACGAAACTGTGAGCTGTGTGCAGAAAGTTGCCCGCTGGTGGCGGG 2647
|||||
DB 715 IleGluIleGlnSerLysTrpLysSerValArgMetValAlaValProArgLys 734
QY 2648 TCCCTGGCGGGGGCCGAGGACGATGAGACCGGAAGAT-GTAAAGCCCATGGGCC 2706
|||||
DB 735 AlaLeuAlaGlyAlaGlnLeu-----GlnGluGluThrGlnThrValSerAlaLeuAla 752
QY 2707 TTGCTGTGTGGGGGGCGAAGACGCGCCAGCTGGCGGACCCCGGACACAGCCCAAG 2766
|||||
DB 753 SerLeuThrValAspValGluGlnProPheAla-----GlnGluAspSerArg 768
QY 2767 CCGGCAAGAGCCCATGTGAGCAGGACCTGCGCTG 2802
|||||
DB 769 ThrAlaGlyGluProMetGluGluGluProAlaLeu 780

RESULT 11
US-10-072-094-107
Sequence 107, Application US/10072094
Publication No. US2003002538A1
GENERAL INFORMATION:
APPLICANT: JACKSON, DONALD
APPLICANT: LORENZI, MATTHEW
APPLICANT: ATTAR, RICARDO
APPLICANT: GOTTARDIS, MARCO
TITLE OF INVENTION: NOVEL HUMAN HISTONE PEPTIDES
FILE REFERENCE: 3053-414US1
CURRENT APPLICATION NUMBER: US/10/072, 094
PRIOR FILING DATE: 2002-06-14
PRIOR APPLICATION NUMBER: 60/298, 296
PRIOR FILING DATE: 2001-06-14
NUMBER OF SEQ ID NOS: 127
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 107
LENGTH: 1141
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Consensus
FEATURE: sequence

NAME/KEY: MOD_RES
LOCATION: (2)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (4)..(7)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (15)..(16)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (21)..(24)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (26)..(27)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (29)..(30)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (35)..(37)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (40)..(41)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (43)..(46)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (49)..(62)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (66)..(67)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (70)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (74)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (82)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (84)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (86)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (95)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES
LOCATION: (104)
OTHER INFORMATION: Unknown consensus marker
FEATURE:
NAME/KEY: MOD_RES

1	LOCATION: (110)..(111)	consensus	marker
2	OTHER INFORMATION: Unknown		
3	FEATURE:		
4	NAME/KEY: MOD_RES		
5	LOCATION: (114)..(115)		
6	OTHER INFORMATION: Unknown	consensus	marker
7	FEATURE:		
8	NAME/KEY: MOD_RES		
9	LOCATION: (119)		
10	OTHER INFORMATION: Unknown	consensus	marker
11	FEATURE:		
12	NAME/KEY: MOD_RES		
13	LOCATION: (126)		
14	OTHER INFORMATION: Unknown	consensus	marker
15	FEATURE:		
16	NAME/KEY: MOD_RES		
17	LOCATION: (128)		
18	OTHER INFORMATION: Unknown	consensus	marker
19	FEATURE:		
20	NAME/KEY: MOD_RES		
21	LOCATION: (134)..(139)		
22	OTHER INFORMATION: Unknown	consensus	marker
23	FEATURE:		
24	NAME/KEY: MOD_RES		
25	LOCATION: (143)..(144)		
26	OTHER INFORMATION: Unknown	consensus	marker
27	FEATURE:		
28	NAME/KEY: MOD_RES		
29	LOCATION: (147)..(148)		
30	OTHER INFORMATION: Unknown	consensus	marker
31	FEATURE:		
32	NAME/KEY: MOD_RES		
33	LOCATION: (158)..(160)		
34	OTHER INFORMATION: Unknown	consensus	marker
35	FEATURE:		
36	NAME/KEY: MOD_RES		
37	LOCATION: (163)		
38	OTHER INFORMATION: Unknown	consensus	marker
39	FEATURE:		
40	NAME/KEY: MOD_RES		
41	LOCATION: (165)..(166)		
42	OTHER INFORMATION: Unknown	consensus	marker
43	FEATURE:		
44	NAME/KEY: MOD_RES		
45	LOCATION: (169)		
46	OTHER INFORMATION: Unknown	consensus	marker
47	FEATURE:		
48	NAME/KEY: MOD_RES		
49	LOCATION: (173)		
50	OTHER INFORMATION: Unknown	consensus	marker
51	FEATURE:		
52	NAME/KEY: MOD_RES		
53	LOCATION: (176)		
54	OTHER INFORMATION: Unknown	consensus	marker
55	FEATURE:		
56	NAME/KEY: MOD_RES		
57	LOCATION: (185)		
58	OTHER INFORMATION: Unknown	consensus	marker
59	FEATURE:		
60	NAME/KEY: MOD_RES		
61	LOCATION: (193)		
62	OTHER INFORMATION: Unknown	consensus	marker
63	FEATURE:		
64	NAME/KEY: MOD_RES		
65	LOCATION: (195)..(202)		
66	OTHER INFORMATION: Unknown	consensus	marker
67	FEATURE:		
68	NAME/KEY: MOD_RES		
69	LOCATION: (205)		
70	OTHER INFORMATION: Unknown	consensus	marker
71	FEATURE:		
72	NAME/KEY: MOD_RES		
73	LOCATION: (208)		

QY	467	GAGATGTCCTCCACACCGACCCCTCCCTCTGAGACGTCCTCCCAACGATTCAAGCTTC	526
Db	349		368
QY	527	TACAGTGTCTCTCTGCGCAACATGTCCTTAGGCTGAGAGCGACAGGCTATGTCACG	586
Db	365		388
QY	587	AACTGACACTTCAGTCCCTCCGGAAGCTGTGACACAGAGAGAGCGCGAGCGACGCC	646
Db	389		408
QY	647	CTCCAGTCCCTCCGGGACAGGTGGACGCTGACCGGCAAGTTCATGACACATCCTATT	706
Db	409		428
QY	707	CGTGGCTGCTGTGGGCGTGCGACATGAGAGGCGACGGAGCCCGACGGCATGCTCC	766
Db	429		448
QY	767	CTGCTGACAGATGTCTGTCTGAGACAGGCCCGGACGACGACCTCATG	820
Db	449		468
QY	821	CGTGGCTGCTGTGGGCGTGCGACATGAGAGGCGACGGGATGAGCTGGGCGACG	874
Db	465		488
QY	875	AGCATGCGGAGCGTAGGACGACCTCCCGGCGATCGGCCCTGACGCGACATGCTCTCA	934
Db	489		508
QY	935	CGGTCGCGGACAGTCCCGGACGCTCTGACAGCTGTCTATGCAACAGCACCGACGAC	994
Db	509		528
QY	995	TTCCTGAGAGACAGAA--	1047
Db	529		548
QY	1043	ACCAAGACAGGGAGCTGCCAGGACGCCACACCCACCCTGAGAGACAGAGAGAGAG	1107
Db	549		568
QY	1103	CTGAGGAGACAGACAGAGCTTCTCTGGGGAGAGAGCCTGACACTGCCCCGGGAGGCG	1165
Db	569		588
QY	1163	TCCACAGAGAGTGTGAGACACACAGAGACCTGTGAGAGAGAGACGAGAGAGATGGG	1222
Db	589		608
QY	1223	GAGAGAGAGAGAGATTGCACTTCAGCTTAAGAGACGAGAGGGCGAG--	1267
Db	609		628
QY	1268	---AATGGTCTGAGAGAGGGGCCGACTTGGAGAGAGCTGTGCTGATACAAAACTG	1322
Db	629		648
QY	1325	TTCATGATGCCCAACCGCTGCACACTTTCAGATGTGACCAAGCGCCCTCAGACCTGGCC	1388
Db	649		668
QY	1385	ACTGTGCCCCACCAAGCCGTGGGCCGTACCAATCTCTCCCTGCTGCCCCCTGGGGGCATG	1444
Db	669		688
QY	1445	AAGAACCCCCAGACCAACCGTATAGACACTCTTCAACCAACAAGTGGTCTTACGACACG	1504
Db	689		708

QY	1505	TTATGCTAAAGACCAAGTGTATGTGGGGAGAACACAGTGCACCCCTAGATGCTGGC	156
Db	709	***MetLeuLysHisGlnCys**CysGly***Ser*****HisProGlnHisAlaGly	728
QY	1565	CGGATCCAGACCACTCTGGTCCGCGCTCCAGAGAACAGAGCCCTGTTAGCAAGTGCAGCGG	1624
Db	729	ArgHisGlnSerHisLeuProSerArgLeuGlnGlnHisGlyLeu*****LysCysGln***	748
QY	1625	ATCCGAGGTGCGAAAGCCACGCTAGATGAGATCCAGACAGTGCATCTGAATATCCACAC	1684
Db	749	His***GlyArgLysAlaSerLeuGlnGlnLeuGln***ValHisSerGlu***HisSer	768
QY	1685	CTGGCTCTTTGGGAGACAGTCCCTCAACGGCCAGACAGTGTAGACAGCAAGAAAGTTGGTGGT	1744
Db	769	LeuLeuTyrGlnTyr***ProLeu*****GlnLysLeuAsp***Arg***LeuLeuGly	788
QY	1745	CCCATCCACCCGAAGATGATGATGTCTGTCTGCTCTTGTGGGGGATCCGGGTGCAGAGTAC	1804
Db	789	*****Phe*****LeuProCysGlyGlyLeuGlnValAlaSerAsp	808
QY	1805	ACCGTGTGAATGAGATGACATCTCTCCAGTGTGTGGCATGGGACGTGGGCTGCTGCTG	1864
Db	809	ThrHisLeuPasnGlnLeuHisSerSer***Ala***ArgMetAlaValGlySerValIle	828
QY	1865	GAGCGGGCTTCAGAGTGGGGCTGCAGAGAGACCTCAAGATGAGATTGCCATATCCGGGCC	1924
Db	829	GlnLeu*****LysValAla***GlyGlnLeuLysAsnGlyPheMetAlaValAlaPro	848
QY	1925	CCAGAGCACACAGCCGCGAGATCCACAGCCATGGAGATTCTGCTTCCAACTCTATACC	1984
Db	849	ProGlnHisHisAlaGlnGlnSerThr***MetGlyPheCysPheMetAsnSerValAla	868
QY	1985	ATCACCCGAAACTCTTACAGACAGAGTGTACAGCTGGGACAGTCTCATCTGTCGATCG	2044
Db	869	Ile***AlaLys***Leu*****Leu***Ile***LysIleLeuIleValAsp***	888
QY	2045	GACATTCACCACTGGCATGGGACACCCACAGGCTTCTACAATGACCCCTGTGTGCTTAC	2104
Db	889	AspValHisHisGlyAsnGlnTyrGlnGlnAlaPheTyr***AspProSerIleLeuTyr	908
QY	2105	ATCTCTGCAATGCCCTTGTGACAACGGGAATCTTCTTCAGGCTCTGGAGCTCTGAAAG	2164
Db	909	IleSerLeuHisArgTyrAsp***GlyAsnPhePheProGlySerGlyAlaPro***Glu	928
QY	2165	GTTCGTGGAGAGCACGAGCGTGGGGGTACATGTGAACGTGGCATGCAAGAGGTGTGAC	2224
Db	929	Val*****Leu***	948
QY	2225	CCCCCATTTGGAAGACGTGGAGTACTTATACAGCTTGTAGAGACAGTGTATGCCATTGCC	2284
Db	949	ProPheMetGlyAsp***GluTyrLeu***AlaPheArgThrIleVal***ProIleAla	968
QY	2285	CACGAGTTCACCTGATGTGGTGTCTAGTGTCCGCGGGGTTTGATGATGCTGTGAAGACAT	2344
Db	969	***GlnPhe***ProAspMetValLeuValSerAlaGlyPheAspIleLeuGlnLysHis	988
QY	2345	CTGTCTCCCTGTGGGTGCTACTCTGTTCACCGCCAGATGTTTGGCCACTTGCACGAGCGAG	2404
Db	989	*****ProLeuGlnGlyGlyTyr***ValThrAlaLysCysPheGlnHisIleThrLysGln	1008
QY	2405	CTGATGACCTTGGGACGGGCGCGGTGTCTGTGGCTCTGTGGAGGAGGCCCATGACTTGAC	2464
Db	1009	LeuMet***LeuAla***GlyArgValValLeuAlaLeuGlnGlyGlyHisAspLeuThr	1028
QY	2465	GGCATCTGTATGCTCTGTAAGCTTGTGTCTGTGGCTCTGTCTAGTGTAAAGCTGCAGCC	2524
Db	1029	AlaIleCysAspAlaSerGlnAlaCysVal***AlaLeuLeu***GlnLeu***Pro	1048
QY	2525	TTGGATGGGAGGACTTGGAGCAAGAAAGCCCAATCAACGACGAGTGGCCATGAGAGAA	2584
Db	1049	Leu***Glu***IleLeu***Gln***ProAsn***AsnAlaVal***SerLeu***Lys	1068
QY	2585	GTCATCAGATCCAGAACAAACCTGAGACTGTGTGCAGAAATTTGCCGCGTGTGTGGC	2644


```

QY 1943 GAATCCAGCCATGGATTCCTGCTTCTTCACTGTACCATCCGCAAAATCTCTA 2002
    |||||
Db 633 HisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysArgGlnLeu 652
QY 2003 CACGACAG------TTGAACGTGGGCAAGGCTCCTCATCTGAGTCGGACATTCAC 2053
    |||||
Db 653 GlnGlnGlnSerLysAlaSerLysAlaSerLysIleLeuIleValAspTrpAspValHis 672
QY 2054 CATGGCAATGGCAACGACGAGCGCTTCTACAAATGACCCCTGTGCTTCAATCTCTG 2113
    |||||
Db 673 HisGlyAsnGlyThrGlnGlnThrPheTyrgLnaAspProSerValLeuTyrlleSerLeu 692
QY 2114 CATGGTATGACAAAGGAACTTCTTCCAGGCTCGGCGCTCCTGTAAGAGGTTGGGGA 2173
    |||||
Db 693 HisArgHisAspAspGlyAsnPhePheProGlySerGlyAlaValAspGluValGlyAla 712
QY 2174 GACACAGGCGGTGGGTACAATGTGAACGTGGCATGGACAGAGGTGTGACCCCCCAT 2233
    |||||
Db 713 GlySerGlyGluGlyPheAsnValAsnValAlaTrpAlaGlyIleLeuAspProPmet 732
QY 2234 GAGAGGTGGAGTACCTTACAGCCTCAGACAGCATGGTGTGCGCATTCGCCACAGCTTC 2293
    |||||
Db 733 GlyAspProGluTrpLeuAlaAlaPheArgIleValValMetProIleAlaArgGluPhe 752
QY 2294 TCACCTGATGTGCTCTAGTCTCCGCGGCTTGAATGCTGTGAAGACATCTGTCTCT 2353
    |||||
Db 753 SerProAspLeuValLeuValSerIleArgIlePheAspAlaIleGluGlyHisProAlaPro 772
QY 2354 CTGGGTGGCTACTCTGTCAACCGCCAGATGTTGGCCACTTGAACGACGAGCTGATGAC 2413
    |||||
Db 773 LeuGlyGlyTyrlleHisValSerAlaIleCysPheGlyTyrlleMetThrGlnGlnLeuMetAsn 792
QY 2414 CTGGCAGGCGCGGCTGCTGCTGCGCCCTGGAGGAGGAGCCATGACTTGAACGCGCATCTGT 2473
    |||||
Db 793 LeuAlaGlyAlaValAlaValLeuAlaLeuGlyGlyHisAspLeuTrpAlaIleCys 812
QY 2474 GATGCTCTGAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2533
    |||||
Db 813 AspAlaSerGluAlaCysValAlaAlaLeuLeuGlyAsnArgValAspProLeuSerGlu 832
QY 2534 GCAGTCTTGACGCAAAAGCC 2554
    |||||
Db 833 GluGlyTrpGlyGlnLysPro 839

RESULT 15
US-09-817-538-13
; Sequence 13, Application US/09817538
; Patent No. US20020137162A1
; GENERAL INFORMATION:
; APPLICANT: LI, Zuomel
; APPLICANT: Bonfils, Claire
; APPLICANT: Besterman, Jeffrey
; TITLE OF INVENTION: Antisense Oligonucleotide Inhibition of Specific Histone
; FILE REFERENCE: 106101.144
; CURRENT APPLICATION NUMBER: US/09/817, 538
; CURRENT FILING DATE: 2001-03-26
; PRIOR APPLICATION NUMBER: US 60/192,157
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 855
; TYPE: PRT
; ORGANISM: Human
US-09-817-538-13

Alignment Scores:
Pred. No.: 7.43e-75 Length: 855
Score: 1671.50 Matches: 384
Percent Similarity: 53.47% Conservative: 101
Best Local Similarity: 42.34% Mismatches: 195

```

```

Query Match: 31.27% Indels: 227
DB: 10 Gaps: 22
US-09-502-945-2 (1-2885) x US-09-817-538-13 (1-855)

QY 11 TTGTCGAAGTCAAAAGAGCCACACAGCGGCGCTCAACATTCCTCCACAGACACCC 70
    |||
Db 101 LeuGluArgThrValHisProAsnSerProGlyIleProIleArgThrLeuGluProIle 120
QY 71 AAATGCTGGGAGCCACCAATGCTTCTTGGACCAAGATTCCCTCCACAGAGGCGCC 130
    |||
Db 121 GluThrGluGlyAlaThrArgSerMetLeuSerSerPheLeuProProValProSerIle 140
QY 131 CGGAGGAGCGCTCCCTCCCTCAAAAGTCCCTTGGCGTGGCGCTACAGCACTCGAGACAG 190
    |||
Db 141 ProSerAspProProGluHis----- 147
QY 191 TTCCTCCCTCCGAAAGAGCCTGTGAACCCCACTTGAAGTGGCTTCAGGCTAAACAG 250
    |||||
Db 148 PheProLeuArgLysThrValSerGluProAsnLeuLysLeuArgTyrlleProLys--- 166
QY 251 AAGTGGCTAGGCGGAGAACAGACATCCCTCTGCTGCGCAAGATGGAGTGTATTAGC 310
    |||
Db 167 LysSerLeuGluArgArgLysAsnProLeuLeuArgLysGluSerAlaPro-----Pro 184
QY 311 ACCTTTACAAGAGAGCCTGTGAGATCACAGGTGCGGCGCTGCGCTGCTGCTGT 370
    |||
Db 185 SerLeuArgArgArgProAlaGluThrLeuGlyAspSerSerProSerSer----- 202
QY 371 AACAGCGACCGCGGCTCCGCGCCAGCTCTCCCAACAGTCCACAGACACCATCGCTGAG 430
    |||
Db 203 ---SerThrProAlaSerGlyCysSerSerProAsnAspSer-----Glu 216
QY 431 AATGGCTTACTGAGTCAATGCCCAACATCCCACTAGATGCTCCTCAGACACGAGCC 490
    |||
Db 217 HisGly----- 218
QY 491 CTCCTCTGACAGACGTCGCCCAACAGTTCAGGCTGTACAGTCTCTCTGCCCCAAC 550
    |||
Db 219 -----ProAsn 220
QY 551 ATCTCCCTAGGCTGACAGGCGACAGGTCACGTACACCAACTCAGACCTCAGCTGCCCG 610
    |||||
Db 221 ProIleLeuGly----- 224
QY 611 AACCTGTGACACAGCAGGAGCGGAGAGGCGCTCCAGTCTCCTGCGGAGGGTGGC 670
    |||
Db 225 -----AspSerAspArgArgThrHisProThrLeuGlyProArgGly 238
QY 671 AGCTGACGCGCAAGTTCATGAGCAGACATCCTCATTCCT----- 709
    |||
Db 239 ProIleLeuGlySerProHisThrProLeuPheLeuProHisGlyLeuGluProGluAla 258
QY 710 ---GCTGCTGCTGCGCGCTGACCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 766
    |||||
Db 259 GlyGlyCysLeu-----ProSer 264
QY 767 CTGCTGACAGATGCTGCTGTGCTGAGACAGGCGCGGAGAGACACCTCATTCCTGTG 826
    |||||
Db 265 ArgLeuGlnProIleLeuLeuAspProSerGlySerHisAlaProLeuLeuThrVal 284
QY 827 CCACTCCAGCGGACGTCGCATGAGCGGAGGAGCGTGGCCACACAGCAGGAGGAGGAG 886
    |||
Db 285 ProGlyLeuGlyProLeuProPheHisPheAlaGlnSerIleMetThrThrGluAlaGlu 304
QY 887 GTAAGCAAGCTCCGCGGCGATCGGCGCTGAGCGGCGGACTGCTCACCGGCTGCGGAG 946
    |||
Db 305 SerGlySer---GlyLeuHisThrProLeuSerArgThrArgSerGluProLeuProPro 323
QY 947 AGTCCCGAGGCG-----CTGACAGCAAGCTGTGATG 976
    |||
Db 324 SerAlaThrAlaProProProProGlyProMetGlnProArgLeuGlnGlnLeu----- 341

```

QY	977	CACACACGACACCGACGTTCTCTGGAGAACACAGACAGACAGCTACAGCTGGGGCAAG	1030
Db	342	--LysThrHisValIleValIleLysArgSer-----	351
QY	1037	ATCTCCACCAAGACAGAGGGAGCTGCTCCAGGACACCCACACACCTGAGGAGACAGAG	1096
Db	352	-----AlaLysProSerGlnLysProArg-----	359
QY	1097	GAGGAGCTGACGAGCAGACAGAGAGCTTCTCTGGGAGAGAGACCTGACATGCCCCGG	1156
Db	359	-----	359
QY	1157	GAGGCTCCACAGAGAGTGGAGAGACACAGAGAGACCTGGAGGAGGAGGAGGAAG	1216
Db	360	-----LeuArgGlnIleProSerAlaGlnLysPheuIleuThr	371
QY	1217	GATGGGAGGAGGAGGAGAGATGTCATCCAGGTAAAGCAGAGAGGGCAGAGTGT---	1273
Db	372	AspGly-----GlyLysProGlnIleu	378
QY	1274	---GCTAGAGAGGGGGCCGACTTGGAGAGACCTGGTGGTATACAAAAACTGTTTCA	1330
Db	379	ValValAspArgLysLeuGlnHisArgGlnLeuGlnHisGlyGln-----Pro	394
QY	1331	GATGGCCCAACCGCTGCACACCTTGACAGCTGACACAGCGCGC-----	1372
Db	395	GlnAlaArgGlyProAlaProLeuGln-----GlnHisProGlnValIleIleThrGln	412
QY	1373	-----CTCAGC	1378
Db	413	GlnGlnArgLeuAlaGlyArgLeuProArgGlySerThrGlyAspLysValIleLeuPro	432
QY	1379	CTGGCCACTGTGCCCCACCAAGCCCTGGCCGTGACCAATCTCCCTGCTGCTCCCTGGG	1438
Db	433	LeuAlaGlnGlyGlnHisArgProLeuSerArgAlaGlnSerSerProAlaIleProAla	452
QY	1439	GGCATGAAGAACCC-----CCACACCAACCC	1465
Db	453	SerIleSerAlaProGlnProAlaSerGlnAlaArgValLeuSerSerSerGlnThrPro	472
QY	1466	GTCACAGCACCTC---TTCACACAGAGTGGTGTACAGACAGCTCAATGCTAAAGCACAG	1522
Db	473	AlaArgThrLeuProPheLeuThrGlnIleLeuLysAspSerValMetLeuLysHisGln	492
QY	1523	TGCATGTGCGGGGAACACACACAGTGCACACCTGACATAGCTGGCGGGATCCAGAGCATCTGG	1582
Db	493	CysSerCysGlyAspAsnSerArgHisProGlnHisAlaGlyIleIleSerIleThr	512
QY	1583	TCCCGGTCAGAGACAGACGCTGCTTGAATGAGGAGCGAGCTCCAGAGTGCACAAACC	1642
Db	513	SerArgLeuGlnGlnAlaArgGlyLeuAlaArgSerGlnCysGlnCysLeuArgGlyArgLysAla	532
QY	1643	ACGCTAGATGATCCAGACAGTGCACCTGTGAATACACACACCTGCTATGAGACAGT	1702
Db	533	SerIleGlnIleuLeuGlnIleuGlnSerValHisSerGlnArgHisValIleuLeuThrGlyThrAsn	552
QY	1703	CCCCCTCAACCGGAGAGCTAGACAGACAGAACTTCTCTGGTCCATCAGCCCAAGATG	1762
Db	553	ProLeuSerArgLeuLysLeuAspAsnGlnLysLeuAlaGlyIleIleLeuAlaArgMet	572
QY	1763	TATCTGTGCTGCTTGTGGGGGCAATCGGGGTGGAGAGTGAACAGCAGCTGTGGAAATGATG	1822
Db	573	PheGlnLeuLeuProCysGlyLysIleGlyValAlaLysPheThrAspThrIleThrPasnIleu	592
QY	1823	CACATCTTCACAGTGTGGCCATGGAGAGTGGCTGCTGGAGCTGGCCCTCAAGGTG	1882
Db	593	HisSerSerAsnAlaAlaArgTrpAlaAlaGlySerValThrAspLeuAlaPheLysVal	612
QY	1883	GCTCCAGAGAGCTCAAGATGATTTGGCATATCCGCGCCCGACAGACACACCCAG	1942
Db	613	AlaSerArgLeuLeuLysAsnGlyPheAlaValAlaValArgProProGlnHisHisAlaAsp	632
QY	1943	GAATTCACAGGCATGGGATTTCTCTTCACTCTGTAGCATCAGCGCAAACTCCTA	2002

Db	633	HisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaIleAlaCysArgGlnLeu	652
QY	2003	CAGCAGAAG-----TTGAACGTGGGCAAGGTCCTCATGCGTGGACTGGACATTGAC	2053
Db	653	GlnGlnGlnSerLysAlaSerLysAlaSerLysIleLeuIleValAspPrpAspAlaHis	672
QY	2054	CATGGCAATGGCACCCAGGAGGGGGTTCACAAATGACCCCTCGTGCTCAACATTCCTGG	2113
Db	673	HisGlyAsnGlyThrGlnGlnThrPheArgGlnAspProSerValLeuArgIleSerLeu	692
QY	2114	CATGCGTATGCAACAGGGAACTCTTCTCCAGGCTCTGGGGCTCTGGAAAGAGGTGGTGA	2173
Db	693	HisArgHisAspAspGlyAsnPhePheProGlySerGlyAlaValAspGluValGlyAla	712
QY	2174	GGACCAGGCGTGGGGTCAATGTGAACGTGGCATGTGACAGAGAGGTGTGACCCGCCAAT	2233
Db	713	GlySerGlyGlnGlyPheAsnValAsnValAlaTrpAlaGlyGlyLeuAspProPheMet	732
QY	2234	GGAGAGCGTAGATCCCTTACAGCCTTCAGAGACAGTGGTATGGATCCCATTCGCCACAGTTC	2293
Db	733	GlyAspProGlyArgLeuAlaAlaPheArgIleValIValMetProIleAlaArgIlePhe	752
QY	2294	TCACCTATGTGGTCCTCAAGTCCCGCGGGTTGAGTGCAGTCGTGAAGAAATGTGCTCT	2353
Db	753	SerProAspLeuValIleValIleSerAlaGlyPheAspAlaIleGluIleIleProAlaPro	772
QY	2354	CTGGGTGGCTACTGTGTACACCGCCAGATGTTGTTGGCCACTTACACAGCAGCGATGTGACC	2413
Db	773	LeuGlyGlyArgHisValSerAlaLysCysPheGlyTyrMetThrGlnGlnIleuMetAsn	792
QY	2414	CTGGCAGGGGCGCGGGTGTGGTGGCCCTGGAGGGAGCCATGACTTGAACCCCATCTGT	2473
Db	793	LeuAlaGlyGlyAlaValIleValIleAlaLeuGlnGlyIleHisAspSerThrAlaIleCys	812
QY	2474	GATACCTCTGAAGCTTGTGCTCGGGCTGTGCTCAGTGTGAATCCGACGCCCTTGATGATAG	2533
Db	813	AspHisSerGluAlaCysValAlaIleAlaLeuLeuGlyAsnArgValAspProLeuSerGlu	832
QY	2534	GCACTCTTGACGACAAAGCCC 2554	
Db	833	GluGlyTyrTrpLysGlnLysPro 839	

Search completed: March 21, 2003, 13:57:15
Job time : 112.891 secs

QY 1943 GAATCCACAGCCATGGATTCCTCTCAACTCTGTAGCCATCACCGCAAACCTCCTA 2002

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 22.35 Seconds

(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346
Sequence: 1 ggaattcccttgcgaagt.....aaacaaagttaaaattt 2885

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_plus_n2p_model -DEV=xlp
-O=/cg2_1/USPTO.spool/US0502945/runcat_14032003_101059_19145/app_query.fasta_1.10979
-DB=Issued_Patents_AA -OFMT=fasta -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -DOCLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptlo -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US0502945 @CGN_1.1.80 @unat_14032003_101059_19145 -NCP=6 -ICPI=3
-NO_XLXLY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- Issued Patents AA:*
- 1: /cg2_6/ptodata/1/iaa/5A_COMB.pep.*
 - 2: /cg2_6/ptodata/1/iaa/5B_COMB.pep.*
 - 3: /cg2_6/ptodata/1/iaa/6A_COMB.pep.*
 - 4: /cg2_6/ptodata/1/iaa/6B_COMB.pep.*
 - 5: /cg2_6/ptodata/1/iaa/6C_COMB.pep.*
 - 6: /cg2_6/ptodata/1/iaa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	374.5	7.0	335	4	US-09-446-504-80
2	374.5	7.0	335	4	US-09-712-266-80
3	302	5.6	482	1	US-08-528-255A-1
4	302	5.6	482	1	US-08-717-365-1
5	278	5.2	517	4	US-09-282-305-8
6	277.5	5.2	488	4	US-09-282-305-2
7	275.5	5.1	2088	4	US-09-548-372D-13
8	275.5	5.1	2088	4	US-09-548-367D-13
9	271.5	5.1	439	4	US-09-282-305-6
10	270.5	5.1	351	4	US-09-282-305-4
11	265	4.9	1400	4	US-08-630-915A-37
12	262.5	4.9	1917	4	US-09-627-650B-5

C 13	262.5	4.9	1917	4	US-09-436-063C-5	Sequence 5, Appl
C 14	260	4.9	432	4	US-09-282-305-10	Sequence 10, Appl
C 15	254	4.7	2601	4	US-09-627-650B-9	Sequence 9, Appl
C 16	254	4.7	2601	4	US-08-436-063C-9	Sequence 9, Appl
C 17	253	4.7	2211	4	US-09-738-884-1	Sequence 1, Appl
C 18	248.5	4.6	2544	4	US-09-627-650B-3	Sequence 1, Appl
C 19	248.5	4.6	2544	4	US-09-436-063C-3	Sequence 3, Appl
C 20	241	4.5	1652	4	US-09-627-650B-1	Sequence 1, Appl
C 21	241	4.5	1652	4	US-09-436-063C-1	Sequence 1, Appl
C 22	239	4.4	2508	4	US-09-627-650B-7	Sequence 7, Appl
C 23	239	4.4	2508	4	US-09-436-063C-7	Sequence 7, Appl
C 24	235	4.4	1958	1	US-07-945-283-2	Sequence 2, Appl
C 25	234	4.3	1958	1	US-07-945-283-2	Sequence 2, Appl
C 26	231.5	4.3	960	4	US-09-219-849-6	Sequence 6, Appl
C 27	229.5	4.3	714	4	US-08-556-978B-61	Sequence 6, Appl
C 28	229.5	4.3	714	4	US-09-247-806-10	Sequence 10, Appl
C 29	227.5	4.3	1065	1	US-08-642-255-12	Sequence 7, Appl
C 30	224.5	4.2	882	4	US-09-413-814-78	Sequence 4, Appl
C 31	224	4.2	720	4	US-09-219-849-4	Sequence 4, Appl
C 32	224	4.2	777	1	US-08-642-255-53	Sequence 53, Appl
C 33	222.5	4.1	1417	4	US-08-900-230-3	Sequence 3, Appl
C 34	222	4.2	960	4	US-09-219-849-5	Sequence 5, Appl
C 35	221.5	4.1	1128	4	US-09-627-650B-11	Sequence 11, Appl
C 36	221.5	4.1	1128	4	US-09-436-063C-11	Sequence 11, Appl
C 37	219.5	4.1	1057	3	US-08-931-820-1	Sequence 1, Appl
C 38	219.5	4.1	1345	2	US-08-977-767-3	Sequence 3, Appl
C 39	217.5	4.1	1461	4	US-09-585-887-9	Sequence 9, Appl
C 40	217.5	4.1	1461	4	US-08-289-578-9	Sequence 9, Appl
C 41	215.5	4.0	762	1	US-08-642-255-120	Sequence 120, App
C 42	215.5	4.0	762	1	US-08-397-633A-31	Sequence 31, Appl
C 43	211	3.9	801	1	US-07-906-349A-6	Sequence 6, Appl
C 44	207	3.9	2088	4	US-09-548-372D-13	Sequence 13, Appl
C 45	207	3.9	2088	4	US-09-548-367D-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1
US-09-446-504-80
Sequence 80, Application US/09446504
Patent No. 6218150
GENERAL INFORMATION:
APPLICANT: UEMORI, Takashi
APPLICANT: SATO, Yoshimi
APPLICANT: FUJITA, Tomoko
APPLICANT: MIYAKE, Kazuo
APPLICANT: MORAI, Hiroyuki
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS
FILE REFERENCE: 1422-408PCT
CURRENT APPLICATION NUMBER: US/09/446, 504
CURRENT FILING DATE: 1999-12-23
PRIOR APPLICATION NUMBER: PCT/JP98/02845
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: JP 9-187496
PRIOR FILING DATE: 1997-06-26
PRIOR APPLICATION NUMBER: JP 9-320692
PRIOR FILING DATE: 1997-11-27
NUMBER OF SEQ ID NOS: 92
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 80
LENGTH: 335
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-09-446-504-80
Alignment Scores:
Pred. No.: 1.27e-17
Score: 374.50
Percent Similarity: 43.98%
Best Local Similarity: 28.27%
Query Match: 7.01%
Length: 335
Matches: 108
Conservative: 60
Mismatches: 121
Indels: 93

[illegible]

Db	307	-----	:::	:::	-----	Lysalalysalalarglu 312
Qy	2549	AAGCCCAACATCAACGACGATGCCACGCTAGAGAAAGTCATCGAGATCCAGACGAACAC	2608			
Db	313	GlutProatgTyrGluTrpIrrLeuLysLeuAlaGluGluValLysAspIleuSerProTrp	332			
Qy	2609	TGGAGC 2614				
Db	333	TrpSer 334				
RESULT 2						
US-09-712-266-80						
Sequence 80, Application US/09712266						
Patent No. 633158						
GENERAL INFORMATION:						
APPLICANT: UEMORI, Takashi						
APPLICANT: SATO, Yoshiaki						
APPLICANT: FUJITA, Tomoko						
APPLICANT: MIYAKE, Kazuo						
APPLICANT: MUKAI, Hiroyuki						
APPLICANT: ASADA, Kiyozo						
APPLICANT: KATO, Ikunoshin						
TITLE OF INVENTION: DNA POLYMERASE-RELATED FACTORS						
FILE REFERENCE: 1422-408PCT						
CURRENT APPLICATION NUMBER: US/09/712,266						
CURRENT FILING DATE: 2000-11-15						
PRIOR APPLICATION NUMBER: US 09/446,504						
PRIOR FILING DATE: 1999-12-23						
PRIOR APPLICATION NUMBER: PCT/JP98/02845						
PRIOR FILING DATE: 1998-06-24						
PRIOR APPLICATION NUMBER: JP 9-187496						
PRIOR FILING DATE: 1997-06-26						
PRIOR APPLICATION NUMBER: JP 9-320692						
PRIOR FILING DATE: 1997-11-27						
NUMBER OF SEQ ID NOS: 92						
SOFTWARE: PatentIn Ver. 2.1						
SEQ ID NO 80						
LENGTH: 335						
TYPE: PRT						
ORGANISM: Pyrococcus furiosus						
US-09-712-266-80						
Alignment Scores:						
Pred. No.: 1,27e-17						
Score: 374.50						
Percent Similarity: 43.98%						
Best Local Similarity: 28.27%						
Query Match: 7.01%						
Gaps: 4						
US-09-502-945-2 (1-2885) x US-09-712-266-80 (1-335)						
Qy	1547	CACCCCTAGCATGCTGGCGGATCCAGAGCATCTGTCGCCGCTCAGAGACACAGCCTG	1606			
Db	20	HisProGluAsnProLysArgLeuGluIleValLeuSerLysValArgGluIleuGlyLeu	39			
Qy	1607	CTTAGCAAGGCGCGGCGGATCCGAGGTCGCAAGCAGCCTGATGATGATCCAGACACTG	1666			
Db	40	GluGluArgIleGluIleGluIleProAsnProValGluGluTrpPhe-----ValGluLysIle	57			
Qy	1667	CACCTCTGAATACGACACACCCCTGCTATGGACCACTCCCTCAGCCGCGAGAAAGCTAGAC	1726			
Db	58	HisAspArgAspTyr-----IleAsnPheValLysGluIle 69				
Qy	1727	AGCAAGAAGTTGCTCGGCTCCCATCAGCCAGAAAGATGATGCTGCTGCTTGGGGGC	1786			
Db	70	ValGluLysGlyIleThrArgLeuAspProAspTrpTyrVal-----83				
Qy	1787	ATCGGGGTGACAGGTAGCAACCGGTGTGAGATGAGATGACATCCTCCAGTGGCTGTGGCATG	1846			
Db	84	-----SerProGlyThrTrp-----SerIleAlaLeuLeu 93				

[illegible]

```

CITY: Kalamazoo
STATE: Michigan
COUNTRY: USA
ZIP: 49008-1699
COMPUTER READABLE FORM:
MEDIUM TYPE: diskette, 3.5 inches, 1.44 Mb storage
COMPUTER: IBM PC/XT/AT Compatible
OPERATING SYSTEM: MS-DOS 5.0
SOFTWARE: WordPerfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/528,255A
FILING DATE: September 14, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP6-227876
FILING DATE: 22-SEPTEMBER-1994
APPLICATION NUMBER: JP7-183763
FILING DATE: 20-JULY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Terryence F. Chapman
REGISTRATION NUMBER: 32549
REFERENCE/DOCKET NUMBER: Furuya Case 1335
TELECOMMUNICATION INFORMATION:
TELEPHONE: (616) 381-1156
TELEFAX: (616) 381-5465
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 482
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
LIBRARY: human fetal lung cDNA library
US-08-528-255A-1

Alignment Scores:
Pred. No.: 1,07e-12 Length: 482
Score: 302.00 Matches: 100
Percent Similarity: 42.66% Conservative: 51
Best Local Similarity: 28.25% Mismatches: 133
Query Match: 5.65% Indels: 70
DB: 1 Gaps: 16

US-09-502-945-2 (1-2885) x US-08-528-255A-1 (1-482)
QY 1547 CACCCTGAGCATGCTGGCGGAGTCAGAGCATCTGGTCCCGGCTGCAGAGACAGCGCTG 1606
   ||||| ||||| ||||| |||||
DB 28 HisPrometLysProHisAlaIleatGmetThrHisAsnLeuIleuAsnIleuAsnIleu 47
QY 1607 CTTAGCAAGTGGCAGGCGGATCCGAGGATCCGAAAGCCAGACCTAGATGAGATCCAGACATG 1666
   ||| ||| ||| ||| ||||| |||||
DB 48 TyrArgLysMetGluIleTyrArgProHisLysAlaAsnAlaGluGluMetThrLysTyr 67
QY 1667 CACACTCT--GAATACACACACCTGCTGTATGGGAGACACAGTCCCTCAAGCGAGAGACTA 1723
   ||||| ||||| ||||| |||||
DB 68 HisSerAspArgTyrIleLysPheLeuAlaArgSerIleArgProAspAsmetSerGluLys 87
QY 1724 GACACGACAGAGTTG-----CTCGGT-----CCCATGACGCGAAGATG 1762
   ||||| ||||| ||||| |||||
DB 88 ---SerLysGlnMetGlnArgPheAsnValGlyGluAspCysProValPheAspIleu 106
QY 1763 -----TATGCTGTGCTGCTTGTGGGGGCGATCCGGGCTGGACAGTGCACACCTGTGGAT 1816
   :::: ||| ||||| ||||| ||| ||| |||
DB 107 PheGluPheCysGlnLeuSerThrGlyGly---SerValAlaSerAlaValLysLeuAsn 125
QY 1817 GAGATGACATCTCCACATGCTGTGGCGCCATGACAGGAGTGGCTGCTGAGACTGACCTTC 1876
   :::: ||||| ||||| ||| |||
DB 126 LysGlnIleThrAspIleAlaValAsnThrPheAlaGlyGlyLeu----- 139
QY 1877 AAGGTGGCTGCAGGAGAGTCAAGAAATGGATTTGCCATCATCCGAGCCCGACAGACACCAC 1936

```

Db	140	-----		141
QY	1937	GCCGAGAAATCCACAGGCATGGATCTGCTTCTTCAACTCTGTACCATCAACCCAAA		1996
Db	142			159
QY	1937	CTCTACACAGCAAGATTGAACGTGTGGCAGGCTCTCATGTGGACGTGGACATTCACCAT		2056
Db	160			179
QY	2057	GCGAATGGCACCCACGACGAGGGCTTCAATAGACCCCTGTGTCTACATCTCTGCAT		2116
Db	180			199
QY	2117	GCGATGACACGAGCACTTCTTCCAGGCTGTGGGCTCTGTGAAGAGGTGTGTGAGAGA		2176
Db	200			217
QY	2177	CCAGGCGTGGGTCTCAATGTGAAAGCTGTGACGACAGAGGTGTGACCCCATTTGGA		2233
Db	218			233
QY	2237	GACGTGAGATACCTTACAGGCTCTCAGAGACAGTGGTATGCCATTCGCCACAGATTCTCA		2296
Db	234	ASPLGUSERTYGIUAIATLIEPHELYSPTOVAIMETSELTYSVALMETIUMETHEGIN		253
QY	2297	CTGTATGTGTCCTAGTCTCCGCCGGGTTTGATGTGTTGAAGACATCTGTCTCTGTG		2356
Db	254	PROSERIALVALIALLAUNGINCYSGLYSERASPERLEUSERGLY-----ASPARGLYU		271
QY	2357	GGTGGCTACCTGTGCACC-----GCCAGATG-----TTGGGCACCTTGACC		2398
Db	272	GLYUSPHEIASLEUHTIRIELYSGLIYNSALALYSCYSVALGLURHEVALYSSERPHE		291
QY	2399	AGGCAGCTGATGACCCCTGCCAGGAGGGCCGGGTGGTCTGCTGCGCTGGAGAGAGCCCATGAC		2458
Db	292	ASPLEUPROMETLEUWETLEUAGLYL-----GLYLYTYRTHR		304
QY	2459	TTG-----ACCGCCATCTGTATGGCTCTGAAAGCTTGTCTCGGGCTGTGCTCAGTGTA		2512
Db	305	IIEARGASVALAIATARGYSATGRTHTYRGIIUTHRALAVALA-----		319
QY	2513	AAGCTGACGCCCTTGATGTAGGACAGCTTGTCACCAAAAGCCC		2554
Db	320	-----LEUSPHTIRGLIUIETPROASNGIULEUPRO		329
RESULT 4				
US-08-717-365-1				
: Sequence 1, Application US/08717365				
: Patent No. 5763182				
: GENERAL INFORMATION:				
: APPLICANT: MAKAMURA, YUSUKE				
: APPLICANT: : FUKUKAWA, YOICHI				
: TITLE OF INVENTION: RPD PROTEIN AND DNA				
: TITLE OF INVENTION: ENCODING THE SAME				
: NUMBER OF SEQUENCES: 2				
: CORRESPONDENCE ADDRESS:				
: ADDRESSEE: FLYNN, THIEL, BOTTELL & TANIS, P.C.				
: STREET: 2026 Rambling Road				
: CITY: Kalamazoo				
: STATE: Michigan				
: COUNTRY: USA				
: ZIP: 49008-1699				
: COMPUTER READABLE FORM:				
: MEDIUM TYPE: Diskette, 3.5 inches, 1.44 Mb storage				
: COMPUTER: IBM PC/XT/AT Compatible				
: OPERATING SYSTEM: MS-DOS 5.0				
: SOFTWARE: Wordperfect 5.0				
: CURRENT APPLICATION DATA:				
: APPLICATION NUMBER: US/08/717,365				
: FILING DATE: 23-SEP-1996				
: CLASSIFICATION: 536				

QY	1547	CAACCTAGCATGCTGGCCGAGATCCAGACAGATCTGGTCCCGGCTGCAGAGACAGAGCCCTG	1606
Db	28	Histromethylspprohisatgllleatgmetltnhisasmeuleuansntyltylleu	47
QY	1607	CTTAGCAAGTCGACGCGAGATCCGAGGCTGCCAAGCCACGCTAGATGAGATCCAGACAGTGG	1666
Db	48	Tyrtatrgylsmetgllleatgtrprohslsyalasnaalagllugllmetltnltyltyr	67
QY	1667	CACCTCT---GAATPACACACCCCTGCTCTATGGACACATGCCCTCAACCGGACAGCTA	1723
Db	68	Histeraspsrpytlrlllelyspheleuaugserlleatgproaspsmetsercltulyr	87
QY	1724	GACGACGAAGATG-----CTCGGT-----CCCATAGCCAGAAGATG	1762
Db	88	---SerlyslmetltnatrgpheasnaValglglauspCysprovalPheaspllyleu	106
QY	1763	---TATGCTGTGCTGCTGTGTGGGCGCATCGGGATGGAGACGTGACACCGTGTGAT	1816
Db	107	Pheglupheycysglneuserthrlyglyl---SerValAlaserAlaVallylleuasn	125
QY	1817	GAGATGACATCCCTCCAGCTGCTGTGGCCATGCGCAGAGTGGGCTGCTGCTGAGACTGGCCTTC	1876
Db	126	LysglnglnthrspilleaValasnltrPalaagllylleu-----	139
QY	1877	AAGGTGCTGCAGAGAGAGCTCAAGATGGATTGGCATTCATCCGCGCCCGGACACACCAC	1936
Db	140	-----HisHis	141
QY	1937	GCCGAGATCCACACCATGGGATTTGCTTTCTTAACCTGTGAGCCATGACCGCAAA	1996
Db	142	ATCTCTACAGCAGAAAGTTGAAGTGGGCAAGGCTTCATCGTGGAGCTGGGAAATTCACCAT	2056
Db	160	lletduclleuleuylsTyrhisglnealryAlleuTyrIleasplleasplleishis	179


```

Db 337 ProGUAsPlyseProvaIasngluTYrGluTYrPheGlyProAspTYrThIeu 356
QY 2549 -----AGCCCAATCAACGAGTGGCCAGCAAGAAATTCATCGATCCAG 2599
Db 357 HIsVaIAlProSerAsmEtGluAsnLYsAsnThrArgGlnGlnIeuAspRIleaG 376
QY 2600 AGCAAA 2605
Db 377 SerLYs 378

RESULT 6
US-09-282-305-2
; Sequence 2, Application US/09282305
; Patent No. 6287843
; GENERAL INFORMATION:
; APPLICANT: Baldwin, Donald A.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
; FILE REFERENCE: 5718-44,
; CURRENT APPLICATION NUMBER: US/09/282,305
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/080,563
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Zea mays
US-09-282-305-2

```

```

Alignment Scores:
Pred. No.: 4,7e-11 Length: 458
Score: 277.50 Matches: 99
Percent Similarity: 44.00% Conservative: 66
Best Local Similarity: 26.40% Mismatches: 143
Query Match: 5.19% Indels: 68
DB: 4 Gaps: 14
US-09-502-945-2 (1-2885) x US-09-282-305-2 (1-458)

```

```

QY 1400 GGCCTGGGCGGAGCCAAATCTCCCTGCTGGCGGCGGAGTGAAGACCCCGAGAC 1459
Db 3 AlaserGlyGluGlyAlaserLeuPro---SerProAlaGlyGlyGluAspRlanIasG 21
QY 1460 CAACCCGTAACACACCTCTTCAACACAACTGTGTGTACAGACACGTTCACTAAAGCAC 1519
Db 22 ArgAlaValSerTYrPheTYrGluProSerIle----- 32
QY 1520 CAGTCGATGTGGGGAACACACAGCTG-----CAACCTGAGCATGTGGCGGATC 1570
Db 33 -----GlyAspTYrTYrTYrGlyGlnGlnIeuIleProMetLYsProIleAsnArgIle 48
QY 1571 CAGACGATGTGTGGCGGCTGCAAGAGACAGGCTCTTGAAGAACTGCGAGCGGATCGA 1630
Db 49 ArgMetAlaIleSerLeuValAlaIleTYrGlyLeuIleAsnArgLeuIleuSerArg 68
QY 1631 GGTTCGAAGGACGCTAGATGATCCAGACAGTGCATGCAATTAACACACCCCTGTC 1690
Db 69 ProTYrProIleSerGluAlaAspRIleArgArgRheIleSerAspTYrValAlaPhe 88
QY 1691 TATGGGACCAAGTCCCTCAACCGGAGAGACTAGACAGCAAGA--GTTGCTGCTGCCAT 1749
Db 89 LeuAlaSerAlaIleThrcLYsAsnProGlyValLeuAspProArgAlaIleLYsArgRheAsn 108
QY 1750 CAGCAGAAAGATGTATGTCTGTGCTGCTGTGGGGGATCGGAGGTGCAGACAGTGA 1809
Db 109 ValGlyGluAsp-----CysProValPhe----- 116
QY 1810 GTGGATAGATGACACTCTCCAGTCTGTGGCATGGAGTGGGCTGCTGTGAGCT 1869
Db 117 -----AspGlyLeuPhePheProPheCysGlnAlaSerAlaGlyGlySerIleGlyAla 133

```

```

QY 1870 GGCCTTCAAGTGTGGCTGCAGAGAGCTCAAGAAATGGATTGGCATTCGCGCCGAGC 1929
Db 134 -AlaValLYsLeuAsnArgLYsAspRIlaAsp-----IleThValAsnTrpAlaG 150
QY 1930 A-----CACACCGCGGAGAAATCCACAGCCATGGGATTTGCTTCTTAACCTGTAGC 1983
Db 150 yGlyLeuIleIleIleIleLYsLeuSerGluAlaSerGlyPheCysTYrValAlaAspRIleVal 170
QY 1984 CATCACCGGAAACCTGTACAGACAGTGAACGTGAACGTGGGAGGCTCATGCTGAGACTG 2043
Db 170 ILeuAla-----IleLeuGlnIleuLeuLYsPheIleAsnArgValLeuTYrValAspRI 188
QY 2044 GGACATTCACATGGCAATGGACACCGAGCGGCTTTCATTAATGAACCCCTGTGCTTA 2103
Db 188 eaSPValIleIleIleIleLYsAspGlyValGlnGlnAlaPhePheThrIleAsnArgValMetTh 208
QY 2104 CATCTCTGTGCATCGCTATGACACAGGAACTTTCTTCCAGGCTGTGGGGCTCCAGAGA 2163
Db 208 rValSerPheIleLYsTYr-----GlyAspPhePheProGlyThrGlyIleIleThAs 226
QY 2164 GGTGTGGAGAGACACAGCGGTGGGTACAAATGAAACGTGGCATGGACAGAGGTGTGA 2223
Db 226 rValGlyAlaAlaGlyGlyLYsIleThValAlaLeuAsnValProLeuSerAspGlyIleAs 246
QY 2224 CCCCCCATTTGGAGACAGTGGAGTACCTTACAGCCTTTCAGACAGATGTGATGCCATTGC 2283
Db 246 P-----AspThrThrPheArgGlyLeuPheGlnCysIleIleLYsValMet 262
QY 2284 CCACAGTTCTACCTGATGTGTCTTACTCTCCCGGGTTGATGCTTGAAGAGACA 2343
Db 262 tGlnValTYrGlnProAspValAlaValLeuGlnCysGlyAlaAspSerLeuAlaGly-- 281
QY 2344 TCTGTCTCTGGGCGGTCTACTGTGTGCACCCGACAGATGTTTGGCCAC----- 2392
Db 282 ----AspArgLeuGlyCysPheAsnLeuSerValLYs-----GlyIleAlaAspCysIle 298
QY 2393 -----TTGACACGAGCAGCTGTATGAACCTGCGAGGCGGCGGTG----- 2432
Db 298 uArgPheLeuArgSerTYrAsnValProMetMetValLeuGlyGlyGly--GlyTYrThI 318
QY 2433 -----TGTGCGCCCTGGAGGAGGCCATG 2456
Db 318 leArgAsnValAlaArgCysTrpCysTYrGluThrAlaVal 331

RESULT 7
US-09-548-372D-13
; Sequence 13, Application US/09548372D
; Patent No. 6420534
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USE
; FILE REFERENCE: 29915/62801
; CURRENT APPLICATION NUMBER: US/09/548, 372D
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 2088
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-372D-13
Alignment Scores:

```

Pred. No.:	1e-10	Length:	2088
Score:	275.50	Matches:	191
Percent Similarity:	31.1%	Conservative:	58
Best local Similarity:	23.88%	Mismatches:	336
Query Match:	5.11%	Indels:	215
DB:	4	Gaps:	37

US-09-502-945-2 (1-2885) x US-09-548-372D-13 (1-2088)

QY	2105	TGTAGACACAGAGGGGCTCTTTGTATGAAAGCCGTGGCTGGGCTTCATTCGATGTGAATCT	204.6
Db	82	CysThrGlyCysThrGlyCys----ThrGlyAlaAlaCysCysCys-----Cys	97
QY	2045	CCGAGTCCACGATGAGACGCTTGCACAGTCACTTCGTGTAGA-----	199.8
Db	98	AlaGlyAlaThrThrThrGlyCysCysAlaThrGlyThrCysThrGlyThrGlyCys	117
QY	1997	-----GTTTGGCGGTATGGCTCAAGATGGTAAGAAGCAAGATC	199.9
Db	118	AlaGlyAlaCysThrGlyAlaAla--CysAlaThrGlyCysAlaCysAlaThrGlyAlaAl	137
QY	1958	CCATGCGGTGTGATTCCTCGCGCTGGCTGTGCTCGGGGCGCGATGATGGCAATTCATCT	189.9
Db	137	aThrGlyThrCysCysAlaGlyAlaAlaThrGlyGlyGlyAlaAlaGlyThrGlyGly	157
QY	1898	TGAGCTTCCTCGACGCCACCTTGAAAGCCACGCTCCAGCAGGACGCCACTGCGATCGCA	183.9
Db	157	yalAthrThrCysAla-----GlyAlaThrCysCysAlaThr	169
QY	1838	CAGACATGGAGAGTGCATTCATTCACAGGTGCATGCTCAGTCCACCCGATGCCCCAC	177.9
Db	169	rCysAlaGlyGly--GlyAlaCysCysAlaAlaAlaAlaCysCysThrGlyCys--	186
QY	1778	AAGCAGACACAGCATACATCTTCCTGGCTGATGGAGCAGCAACTTCCTGCTACTACT	171.9
Db	187	-----AlaThrThrGly-----AlaThrAlaCysCysAlaAla	197
QY	1718	TCTGCCGCTTGAAGGCGACTGCTCCCATPAGACAGGCTGTGT--ATTTCAGATGCACTG	166.2
Db	198	GlyGlyAlaAlaGly-----GlyCysAlaThrCysCysThrGlyCys	211
QY	1661	TCTGATCATCTAGCGTGGCTTTTCGACATCGATTCGCTCGCACTTCG-----	161.1
Db	212	AlaGlyThrAlaThrThrGlyCysCysAlaAlaGlyAlaAlaGlyThrCysThrAlaCys	231
QY	1610	-----TAACGAGCGCTGTCTCTCGACGCGGAGCAGATGC-----	157.5
Db	232	CysCysThrGlyAlaAlaAlaCysThrGlyCysAlaGlyAlaThrCysAlaCysCysAlaAla	251
QY	1574	TCTGATCGCGCCAGCATGCTCTCAGGGTGCACGTGTGTTCCCGACATGC-----	153.4
Db	252	ThrGlyThrGlyGlyThrAlaGlyAlaAlaGlyCysCysAlaAlaCysCysAlaAlaCys	271
QY	1533	-----ACTGCTGCTTACATGACATGACGTGCTGACACCACTTGTGTGAAGAGT	147.3
Db	272	CysAlaGlyThrGlyAlaCysCysAlaThrCys-----Cys	283
QY	1472	GCTTACGAGGTGTGCTGGGGGCTTCTTCATGACCCCGACAGGCGAGGAGA-----	141.8
Db	284	AlaGlyAlaAlaCysThrGlyGlyThrGlyCys--AlaAlaGlyCysGlyGlyGlyGlyCys	303
QY	1417	TTTGGTAGCGCCAGGCGTTGGTGGGGACACATGGCCAGCTGAGGGCGCT-----	136.6
Db	303	scYsglyCysAlaAlaGlyCysAlaGly--ThrGlyCysAla--AlaGlyAlaCysCysC	322
QY	1365	-----T	136.5
Db	322	ysAlaThrCysCysCysAlaCysThrThrThrGlyThrGlyAlaThrThrCysCysC	342
QY	1364	GGTACACCTGCACAAAGTTGACGGGCTTGGGCACTGTGAGAACGTTTTTGTATCCAGAC	130.5
Db	342	ysThrAlaCysCysGlyCysThrGlyCysThrThr-----	353

[illegible]


```

Db 446 lacys---ThrGlyGlyCysAlaCysAlaCysGlyThrCysGlyCysCysAlaAla 465
QY 953 GGGAGCTCTGGCGGCGGAGGAGTGC---GCTCAGAGGCGCATGGCGCGGA 897
Db 465 laGly---AlaGly-AlaCysAlaThrGlyCysAlaGlyThrGlyAlaAlaGly 483
QY 896 GCT-----TGCTACCGTCGCTGCTG 870
Db 484 AlaGlyThrAlaCysAlaAlaCysThrThrGlyCysAlaThrGlyAlaCys----- 501
QY 869 CCACAGCTTCACCCGCTACTAGTGGGAGTCCCGTGGAGTGCACAGCAATGAGGTGC 810
Db 502 -----ThrAlaCysGlyGlyCysAlaThrThrGlyThrGlyCysThrGlyCys 517
QY 809 TCTGCTGCCGG-----GGCTGCTCCAGCAAGCAATCAT 777
Db 518 --CysCysThrGlyCysGlyGlyAlaAlaThrThrGlyAlaCysAlaAlaGlyThrC 537
QY 776 GCTGCAGCAGGAGGAGCATGCCGCGGCGCTCCGCTCCAGTGCACGCCCGCA 717
Db 537 yscysGlyAlaGlyGly-----GlyGlyThrAlaGlyAlaGlyThrThrGly 554
QY 716 GGCAGCGCAGGAATAGAGATGTCTATGACTTGGCGGTACCGTCCACCTGCCGA 657
Db 554 hrglyThrGlyThrThrGly-CysCysCysAlaCysThrGlyGlyCysThrGlyAlaAla 573
QY 656 -----GGGACTGGAGGCGCTGGCTGCTCGG 633
Db 574 GlyAlaAlaGlyAlaGlyThrGlyAlaCysAlaAlaThrGlyThrGlyAlaThrCys 593
QY 632 CCTCCTGCTGTG---TCGACAGCTTCGGGAGGAGCAGTGAAGTGTGATGATGTA 576
Db 594 ThrGlyCysThrGlyAlaThrGlyCysGlyGlyAlaGlyAlaGly----- 609
QY 575 CCGTGGCTGCACCCCTAGG-----AGATGTTGGCAGAGAGAGAGAGCTGACAGC 522
Db 610 -----GlyAlaThrGlyAlaCysThrCysGlyGlyAlaThrGlyThrCysThrGly 626
QY 521 TGAAGTGTGGGAGGAGTCCAGAGGAGG----- 499
Db 627 GlyThrGlyGlyGlyGlyGlyGlyGlyAlaGlyCysAlaGlyAlaCysAlaCysAlaGly 646
QY 488 -----CTCGTGTGAGGAGGAGCATCTCAGTGGAGTGGGAGGAGGAGCAG 441
Db 647 AlaCysThrAlaThrGlyCysAlaGlyAlaThr-GlyGlyGlyAlaGlyThrGlyAlaAl 666
QY 440 TAAAGCATTCACGAGGAGTGTGAGAGTGTGGAGAGCTGGGCGGAGCCGG 381
Db 666 aglyAlaCysAlaAlaAlaGlyThrAlaGlyThrAlaGlyAlaAlaGlyThr--AlaGly 685
QY 380 GTGCGGTTCACACGAGGAGGAGCCCGGAGCCGCTGATGATCTCAAGCTCTCT 321
Db 685 yCysAlaGlyAlaGlyAlaGlyAlaGlyAlaAlaGlyThrGlyGly----- 703
QY 320 TCTTAAAGTGTAAATACAGTCCATCTTGGAGCAGGAGGAGGAGCTCTCGCT 261
Db 704 -CysThrGlyAlaGlyGlyThrGlyAlaAlaGlyAlaAlaGlyAlaAla----- 720
QY 260 CAGCCACCTTCTGTGTTAGCTTGAACGCACTTCAAGTTGGTTCAGAGGCTGTTGC 201
Db 721 -----GlyAlaAlaGlyCys--Cys 726
QY 200 GAGAGG---GGAAGTGTCTCGACTGCTGTAAGGCCAGGCAAGGAGTTGTAGAGG 144
Db 727 GlyAlaThrGlyAlaThrGlyAlaCys-----GlyAlaGlyAlaCysGlyAla 743
QY 143 GAGGCTCCAGGAGGCGCTGCGGAGGAGGAGCACTGTCCAAAGAGCATGTGG 84
Db 744 ThrGlyAlaGlyAlaThrGlyAlaGlyThrGlyAlaThrGlyAla-----GlyGly 760
QY 83 CTCGCCAGCATTTGGGCTGCTGGAGGAGGAATGT 48
Db 761 ThrAlaGlyAlaGlyAlaAlaGlyAlaGly 772

```

```

RESULT 9
US-09-282-305-6
; Sequence 6, Application US/09282305
; Patent No. 6287843
; GENERAL INFORMATION:
; APPLICANT: Baldwin, Donald A.
; APPLICANT: Briggs, Steven P.
; APPLICANT: Crane, Virginia C.
; TITLE OF INVENTION: Maize Histone Deacetylases And Their Uses
; FILE REFERENCE: 5718-44,
; CURRENT APPLICATION NUMBER: US/09/282,305
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/080,563
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 439
; TYPE: PRF
; ORGANISM: Zea mays
US-09-282-305-6

```

```

Alignment Scores:
Pred. No.: 1,18e-10 Length: 439
Score: 271.50 Matches: 90
Percent Similarity: 42.94% Conservative: 59
Best Local Similarity: 25.94% Mismatches: 107
Query Match: 5.08% Indels: 91
DB: Gaps: 16
US-09-502-945-2 (1-2885) x US-09-282-305-6 (1-439)

```

```

QY 1547 CACCCATGAGCATGCTGGCCGAGATCCAGAGCATGTGGTCCGCGCAGAGAGGCTG 1606
Db 41 HisProMetLysProHisArgIleArgMetThrHisAlaLeuGluGlyArgGlyLeu 60
QY 1607 CTTAGCAAGTGCAGGCGGATCCGAGTCCGAGCCAGCCAGCTAGATGATCCAGAGCTG 1666
Db 61 LeuAspIleMetGlnValPheArgProHisProIleArgAspArgLeuCysAlaPhe 80
QY 1667 CACTCT---GATATACACACCTCTCTATGGAGCAGTCCCTCAACCGGAGAACTTA 1723
Db 81 HisAlaAspArgPyrValSerPheLeuArgSerValThrProGluThrGlnGln----- 98
QY 1724 GACAGCAGAGATTTGCTCGT-----CCATCAGCCAGAG 1759
Db 99 AspGlnIleArgAlaLeuLysArgPheAsnValGlyGluAspCysProValPheAspGly 118
QY 1760 ATGATGCTGTGCTGCTGCT-----GGGGGAGTGGGGGAGGAGTGAACCGCTG 1810
Db 119 LeuTyrSerPheCysGlnThrTyraAlaGlyLysValGly----- 132
QY 1811 TGGAAATGAGATGCATCTCTCAGTGTGTGGCAGTGGAGTGGGCTGCTGAGCTG 1870
Db 133 -----GlyAlaValLysLeuAsnHisGly-----HisAspIle 143
QY 1871 GCTTCAAGTGGCTGCAGAGAGCTCAGAGATGATTTCCATATCGGCCCCAGGA 1930
Db 144 AlaIleAsnThrPheAlaGlyLeu----- 151
QY 1931 CAGCCCGGAGGAGAAATCCAGAGCATGGAGTTCGTTCTTCACTCTGTAAGCATCAC 1990
Db 152 HisHisLalLysLysCysGluAlaSerGlyPheCysTyrValAsnAspIleValAla 171
QY 1991 GCAAAACTCTACAGCAGAGATGAACTGAGCTGGAGAGTCTCATCTGAGTGGACAT 2050
Db 172 -----IleLeuIleuLeuLysTyrGlnHisGlnArgValLeuTyrValAspIle 189
QY 2051 CACCATGGCAATGGACACCCAGAGGCTTCTACATGACCCCTGTGCTATCATCTCT 2110
Db 190 HisHisGlyAspLeuValGluGluAlaPheTyrThrAspArgValMetThrValSer 209

```


Db	900	-----	AlaGlyTThrCysCysGlyThrGlyThrCysCysAlaCys	913
Qy	2645	-----GGCCCAAGACACAGCGGCACTTTCACACACAGCTCCAGGTGTTCCTGGATCT	259	
Db	914	GlyAlaGlyThrCysThrThrAlaThrGlyAlaGlyThrThrGlyCysCysGly--CysA	933	
Qy	2591	CGATGACTTTCCTCAAGCGTGCGCCACTGCGTGAATGTTGGCTTTTGCTGCACAGCTGCC	253	
Db	933	laGlyThrThrThrThrAlaCysThrThrCysAlaGlyThrCysThrAlaThrCysA	953	
Qy	2531	CATCCACAGGCG-----TGACGCTTTCACACTGACGACAGCGGAGACACAGCTT	248	
Db	933	laAlaGlyGlyThrCysGlyThrCysAlaAlaThrCysAlaThrAlaCysGlyCysAla	973	
Qy	2483	CAGAGGCACTCAGATGGCGGTCAAGCAATGCGCTCCCTCAGAGCGACGACCCCGGC	242	
Db	973	laAlaAlaGlyCysThrThrThrAlaGlyThrThrCysAlaGlyGlyAlaGlyAlaAlaThrA	993	
Qy	2423	CCCCGCCAGGGTATCATGCTGCTGCTGCTCAAGTGGCCAAACATCTGGCGTGACAGACT	236	
Db	993	laThrThrCysCysCysGlyCys-----	100	
Qy	2363	AGCCACCCAGAGAGACAGATGCTCCTCAACAGCATAAACCCGGCGAGACTAGAGACA	230	
Db	1001	-----CysThrThrThrGlyCysThrGlyGlyThrThrCysThrT	101	
Qy	2303	CA---TCAGTGAAGAACTCGCGGCAATGGGCAATCACACACTGTCGAAAGCGCTGAAGGT	224	
Db	1014	hrcysCysThrAlaThrThrCysAlaAlaAlaGlyCysGlyThrAla	102	
Qy	2246	ACTCCAGCTCTCCAAATGGGGGGGTCCACA-----	221	
Db	1029	---AlaCysAlaThrCysGlyGlyCysThrThrCysThrAlaCysAlaThrCysAlaThrC	104	
Qy	2217	-----CCCTCTGTCCATGCGCCACTTCACATTTGAACCCAGCGCTGGTCTCCACCACT	216	
Db	1048	ysCysAlaAlaAlaAlaThrAlaThrAlaThr--CysThr--	105	
Qy	2162	CTTCAGAGCCCCCAGAGCTGGAAGAAGTTCGCGTGTGTCATGCGATGACAGAGATGT	210	
Db	1060	-----AlaCysCysAlaThrCysThrGlyThrCys	106	
Qy	2102	-----AGAGCAGAGAGGGTCATTTGTAGAACGCGCTGGTGCCATTGC-----	205	
Db	1070	CysThrGlyAlaThrThrThrGlyThrCysGlyTThrCysAlaThrCysThrCysAlaThrGly	108	
Qy	2057	---CATGGTGAATGCTCCACGTCACGATGAGAGACTTGCCACGTTCACTTGTCT	200	
Db	1090	GlyGlyThrAlaThrCysThrThrThrThrGly-----GlyThrThrGly	110	
Qy	2003	GTAAGAGTTTGGCGTGAATGGCTTCACAGTTGAAGAAGCAGAAATCCCATGGCGTGGATG	194	
Db	1106	AlaGlyCys--CysGlyCysGly-----AlaThrGlyCysGly-A	111	
Qy	1943	CTCGGCGGTGGTGTCTGGGGGGCGGATGATGCCAAATCCATTCTTGAGCTCTCTCCAG	188	
Db	1118	laCysAlaCysCysGlyGlyCysAlaAlaGlyAla-----GlyThrThrGlyC	113	
Qy	1883	CCACCTTGAAAGGCCAGCTCCAGCAGGAGCCCACTGCAATGGCAGACAGCATGGAGAGT	182	
Db	1134	ysThrCysThrCysGlyGly--AlaGlyThrCysAlaCysCys-----	114	
Qy	1823	GCATCTCATTCACACAGGTGCACTGCCA-----	179	
Db	1148	AlaCysThrGlyThrGlyCysThrCysAlaCysAlaAlaThrGlyAlaCysThrAlaCys	116	
Qy	1793	-----CCCCGATGCCCCACAGGCGACAGCAGACAGATCACTCTT	175	
Db	1168	ThrThrThrGlyAlaThrGlyAlaCysCysAlaThrGlyAlaCysThrAlaAlaThrA	118	
Qy	1754	GGCTGATGGGACCGAGCACTTCTGTGCTGTAGCTTCTGCGGTGAAGGAGCATGGTCC	169	

[illegible]


```
Db 1436 CysAlaCysAlaThrAlaCysGlyCysThrAlaThrGly-----GlyCysCysGly 1452
QY 1059 AGCTCCCTGCTTGTGGATGATCTGCCAGCTGATGCTGC----- 1015
Db 1453 -----GlyCysThrCysCysAlaThrThrCysGly 1462
QY 1014 TGCCTTGC-----TTCTCCAGGAACCTGCTGCTGTGT 979
Db 1463 CysGlyCysGlyGlyCysCysGlyAlaAlaAlaAlaGlyCysThrThrCysCys 1482
QY 978 TGC-----ATGACCAGCTGCTGCAGGCGCTGGGACTC----- 946
Db 1483 AlaAlaGlyAlaCysAlaThrGlyCysThrGlyCysCysAlaAlaCysGlyAlaThrGly 1502
QY 945 -----TGGCGCAGCGGTGAGACTGATGCGGC----- 918
Db 1503 GlyAlaCysGlyCysCysThrGlyCys-AlaAlaAlaAlaAlaThrCysGlyAlaThrAl 1522
QY 917 -TCAGGGCGCATGCGCGAGCGCTGCTACGCTCCGATGCTGGCCACACGTTCA 859
Db 1522 aAlaGlyCysThrThrAlaGlyCysGlyAlaThrAlaCys-GlyGly-Thr----- 1539
QY 858 CCGCTCACTAGTGGGAGCGCGCGCTGCGACAGCAATGAGGCTGCTGCGCGG 799
Db 1540 -----ThrThrCysCysAlaThrThrGlyThrCysThrThrCysThr 1557
QY 798 GCC-----TCCTCCAGCACAAGCACAATGCTGCAGCAGGAGGATGCCGCTGG 751
Db 1558 AlaThrCysThrThrCysAlaAlaThrAlaThrAlaGlyThrCysThrAlaCysThrGly 1577
QY 750 GGGCTCCCTCGCCCTCAGTGCAGCCAGCCAGCAGCAGGATGAGATGCTGCTC 691
Db 1578 GlyThrThrGlyThrAlaThrAlaThr-GlyAlaAlaAlaThrAla----- 1592
QY 690 ATGAATCTGCCGCTCAGCG---TGCCACCTGCCGCA-----GGGACTGGAGG 646
Db 1593 -----ThrCysThrAlaAlaGlyCysThrThrAlaAlaAlaCysThrCysGlyThrCysG 1611
QY 645 GCCGCTCTCGCGCTCTGCTGTGTCGACAGCTTCGGGGAGGAGCATGAGGTGTGAGTTG 586
Db 1611 yGlyAlaCysAlaAlaGlyAlaThrCysCysAlaGlyGlyAla----- 1625
QY 585 GTGACAGTGAACCGTGGCTGCAGCCCTAGGAGATGTTGGCAGAGAAGAGACGTGTAG 526
Db 1626 -----GlyAlaAlaCysGlyAlaCysAlaAlaGlyThrGlyGlyCysAl 1640
QY 525 AGGCTGAACGTGGTGGGGAGCTGTCCA---GAGGAGAGGCTCGCTGAGGAGCATC 469
Db 1640 aGlyCysAlaGlyAlaThrCysCysAlaCysThrGlyAlaThrGlyCysGlyThrAlaThr 1660
QY 468 TCACTGGGAGTGTGGGAGCTAGACCAATAAGCCATTCTCAGCGATGTGCTGGGAG 409
Db 1660 rThrCysGlyAlaCysGly-----GlyCysCys-GlyAla 1671
QY 408 CTGTTGGAGAGCTGGGGCGAGCGGCTGC---GCTGTATCACAGCAGCAGCAGCCCA 352
Db 1671 laAlaAlaThrCysGlyAlaGlyThrAlaCysAlaAlaAlaThrGlyGlyThrGlyThra 1691
QY 351 GGGCCGGGACACCTGT 338
Db 1691 lacysglythrCys 1695
```

Search completed: March 21, 2003, 13:07:56
Job time : 71.35 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 08:38:27 ; Search time 70.4665 Seconds

(Without alignments)
10910.938 Million cell updates/sec

Title: US-09-502-945-2

Perfect score: 5346

Sequence: 1 ggaattcccttctgtcgaagt.....aaacaaaagttaaaattt 2885

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+np.model -DEV=xlp
-Q=/cgn2_1/USPPO.spool/US09502945/runat_14032003_101057_19087/app_query.fasta.1.10979
-DB=A.Geneseq.101002 -QEMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LIST=45 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTEMT=ptc -THR_SCORE=Pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-USRP=US09502945 -CGN_1.1.365 -runat_14032003_101057_19087 -NCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEOUTRT -NES_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq.101002:*
1: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1983.DAT:*
4: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1984.DAT:*
5: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1985.DAT:*
6: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1986.DAT:*
7: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1987.DAT:*
8: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1988.DAT:*
9: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1989.DAT:*
10: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1990.DAT:*
11: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1991.DAT:*
12: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1992.DAT:*
13: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1993.DAT:*
14: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1994.DAT:*
15: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1995.DAT:*
16: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1996.DAT:*
17: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1997.DAT:*
18: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1998.DAT:*
19: /SIDS2/gcgdata/geneseq/genesep-emb1/AA1999.DAT:*
20: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2000.DAT:*
21: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2001.DAT:*
22: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2002.DAT:*
23: /SIDS2/gcgdata/geneseq/genesep-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4798	89.7	1141	21 AAB43008	Human ORFX ORF2772
2	4669	87.3	897	20 AA07092	Colon cancer assoc
3	3659	68.4	716	22 AAB49958	Human histone deac
4	3612.5	67.6	717	23 AA099660	Human histone deac
5	2691	50.3	967	22 AAB49957	Human histone deac
6	2685	50.2	967	22 AA099659	Human histone deac
7	2127	39.8	417	22 AA093526	Human polypeptide,
8	1831.5	34.3	1008	22 AA078891	Human protein SEQ
9	1831.5	34.3	1020	22 AA079875	Human protein SEQ
10	1739.5	32.5	1428	21 AA079023	Caspase 8-Interact
11	1671.5	31.3	855	21 AA099662	Human Histone deac
12	1607.5	30.1	614	22 AAB92700	Human protein sequ
13	1601.5	30.0	1235	22 AAB58493	Drosophila melanog
14	1573	29.4	574	21 AA097032	Caspase 8-Interact
15	1552.5	29.0	541	21 AAB42845	Human ORFX ORF2609
16	1484.5	27.8	425	23 ABB89750	Human polypeptide
17	1296	24.2	328	21 AAB45473	Human secreted pro
18	1103.5	20.6	327	21 AAB45474	Human secreted pro
19	945	17.7	191	22 AAB94542	Human protein sequ
20	911.5	17.1	276	22 AAB92577	Human protein sequ
21	878.5	16.4	361	22 AA023362	Novel human enzyme
22	771.5	14.4	1215	23 AAE22856	Human histone deac
23	767.5	14.4	1215	23 AA099661	Human histone deac
24	741.5	13.9	726	22 AAB93058	Human protein sequ
25	733.5	13.7	883	22 AAB62141	Drosophila melanog
26	708	13.2	552	21 AAG51590	Arabidopsis thaila
27	704	13.2	499	21 AAG51591	Arabidopsis thaila
28	702	13.1	468	21 AAG51592	Arabidopsis thaila
29	700	13.1	552	21 AAG09530	Arabidopsis thaila
30	696	13.0	499	21 AAG09531	Arabidopsis thaila
31	694	13.0	468	21 AAG09532	Arabidopsis thaila
32	645.5	12.1	817	22 AAB95640	Human protein sequ
33	623.5	11.7	662	23 ABB81614	Human HDAC9 vari
34	623.5	11.7	673	23 ABB81613	Human HDAC9 prot
35	623.5	11.7	673	23 AAE22854	Human histone deac
36	623.5	11.7	673	21 AAB42743	Human ORFX ORF2507
37	623.5	11.7	686	21 AAB42757	Human ORFX ORF2521
38	580.5	10.9	330	22 AAB23410	Novel human enzyme
39	574.5	10.7	625	23 ABB81615	Human HDAC9 vari
40	527	9.9	363	22 AA000244	Human histone deac
41	509.5	9.5	184	22 AAB94652	Human protein sequ
42	465	8.7	164	21 AAB43314	Human ORFX ORF3078
43	464.5	8.7	170	22 AAB95880	Human protein deac
44	445	8.3	112	22 AAB49959	Human histone deac
45	374.5	7.0	335	20 AA097110	Thermostable poly

ALIGNMENTS

RESULT 1
AAB43008 standard; Protein: 1141 AA.
AAB43008:
08-FEB-2001 (first entry)

Human: open reading frame; ORFX; detection; cytosolic; hepatotropic;
vulnerable; antiproliferative; antiparkinsonian; neurotrophic; neuroprotective;
anticonvulsant; osteopontin; antitumor; immunosuppressant; cardiant;
immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
antianemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;

KM cholesterol ester storage; systemic lupus erythematosus; infection;
KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
KM thrombosis; contraceptive.
XX
OS Homo sapiens.
PN W0200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinketsu RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR N-PSDB; AAC7217.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11: Page 4728-4731; 5507PP; English.
XX
XX
CC AAC74446 to AAC7606 encode the proteins given in ABA40237 to ABA43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
CC antiproliferative; antiparkinsonian; nootropic; neuroprotective;
CC osteoplastic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; antitumor; antileukemic;
CC antineoplastic; antibacterial; antiviral; antifungal; antihelminthic;
CC antihypertensive; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1141 AA:

Alignment Scores:
Pred. No.: 3,82e-315 Length: 1141
Score: 4798.00 Matches: 927
Percent Similarity: 99.57% Conservative: 3
Best Local Similarity: 99.25% Mismatches: 4
Query Match: 89.75% Indels: 1
DB: 21 Gaps: 0

US-09-502-945-2 (1-2885) x AAB43008 (1-1141)
QY 2 GAATTCCTCTTGTGCAAGTCAAGAGCCACGAGGCGCTCAACCATTCCTCCCA 61
DB 208 GlnPheLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAnsHisSerLeuPro 227
QY 62 CAGCAGCCCAAAATGCTGGGAGCCGACATGCTTTGGACGAGGTTCCCTCCCGAG 121
DB 228 GlnHisProLysCysTrpGlyAlaHisHisAlaSerLeuAnsPcInSerSerProGln 247

QY 122 AGCGGCCCCCTGGAGCGCTCCCTCTACAACTGCTTTGGCTGGCCCTACGACAGT 181
DB 248 SerGlyProProGlyThrProProSerTyrLysLeuProLeuProGlyProTyrAspSer 267
QY 182 CGAGACGACTTCCCTCCGCAAAACAGCCTCGAACCCACTTGAAGTGGCTTCAAG 241
DB 268 ArgAspAspPheProLeuAlaGlyThrAlaSerGlnProAnsLeuLysValaGSerArg 287
QY 242 CTAAGACAGAGGTGGCTGGAGAGAGAGGCTCCCTGGCTGGCAAGATGGAGAGT 301
DB 288 LeuLysGlnLysValaGlnAlaArgArgSerProLeuLeuArgArgLysAspGlyThr 307
QY 302 GTTATTAGACCTTTAAGAAGAGAGCTGTGAGATCAAGGTCCGCGCTGGGCGCTG 361
DB 308 ValIleSerThrPheLysLysArgAlaValGlnIleThrGlyAlaGlyProGlyAlaSer 327
QY 362 TCCGTGTATACAGCGACCGGCTCCGGCCCGACGCTTCCCAACAGCTCCACAGCAC 421
DB 328 SerValLysAnsSerAlaProGlySerGlyProSerSerProAnsSerHisSerThr 347
QY 422 ATGCGTGAAGATGCTTACTGCTCAGTCCAGTCCCATATCCCAATGCTCCCTCCAG 481
DB 348 IleAlaGlnAnsGlyPheThrGlySerValProAnsIleProThrGlnLeuLeuProGln 367
QY 482 CACGAGCCCTCCCTCGACAGCTCCCGCAACAGCTTGAAGCTTCAAGTCCCTTCT 541
DB 368 HisArgAlaLeuProLeuAnsPserSerProAnsGlnPheSerLeuThrProSer 387
QY 542 CTGCCCAACATCTCCCTAGGCTGCGAGCGCCAGCTGACGTCAACATCACACTCACT 601
DB 388 LeuProAnsIleSerLeuGlnAlaThrValThrValThrAnsSerHisLeuThr 407
QY 602 GCTCCCGAAGGTGTCGACACAGCAGAGGCGGAGAGGAGGAGGAGGAGGAGGAGG 661
DB 408 AlaSerProLysLeuSerThrGlnGlnGlnAlaGlnAlaArgGlnAlaLeuGlnSerLeuArg 427
QY 662 CAGGTGGCAGGCTGACCGGCAAGTTCATGAGACATCTCTATTCCTGGAGCTGCTG 721
DB 428 GlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerSerIleProGlyCysLeuLeu 447
QY 722 GCGCTGGCACTGGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 781
DB 448 GlyValAlaLeuGlnAlaArgGlnGlnSerProHisGlyHisAlaSerLeuGlnHisVal 467
QY 782 CTGTTCTGGAGAGGAGGCGGAGCAGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 841
DB 468 LeuLeuLeuGlnGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 487
QY 842 TCCCACTAGTGAAGGAGTGAAGCTGTGGCCACACAGCATGGAGGAGGAGTCCCG 901
DB 488 SerProLeuValThrGlnGlnAlaArgValAlaThrSerMetArgThrValGlyLysLeuPro 507
QY 902 CCGCATCGGCGCTGAGCGGCACTCAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAG 961
DB 508 ArgHisArgProLeuSerArgThrGlnGlnSerProLeuProGlnSerProGlnAlaLeu 527
QY 962 CAGCAGCTGCTCATGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1021
DB 528 GlnGlnLeuValMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 547
QY 1022 CTACAGCTGGGCAAGATCTCTACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1081
DB 548 LeuGlnLeuGlnLysIleLeuThrLysThrGlnGlnLeuProArgGlnProThrThrHis 567
QY 1082 CCTGAGGAGACAGAGAGAGAGAGTGAAGGAGCAGCAGCAGCAGCAGCAGCAGCAGC 1141
DB 568 ProGlnGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 587
QY 1142 CTGACCATGCGCGGAGGAGGCTCCACAGAGAGTGAAGCAGCAGCAGCAGCAGCAGCAG 1201
DB 588 LeuThrMetProArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 607

QY	1202	GAGGACGAGGAAAGAGANTGGGAGAGAGAGAGAGAGATTCATCCAGGTTAAGACAGAGAG	1261
Db	608	GUAspRGUGUGUGUAspRGUGUGUGUGUGUGUAspCysIleGlnValIAspPdlu	627
QY	1262	GCCGAGATCGGTGTGAGAGAGAGGGCCGAGTTGTGGAGAGAGCTGTGGTGCATGACAAAAA	1321
Db	628	GIyGIUsRGUyAlaIGUGUGUGUyPProAspRGUGUGUGUGUGUGUyAlaIGUyIyUs	647
QY	1322	CTGTCTCAGATGCCCAACGGCTGCACACTTTGTACGGTGTACCAAGCGGCCCTCAGCCGTG	1381
Db	648	LeuPheSerAspAlaIGInProLeuGInProLeuGInValIyGlnAlaProLeuSerLeu	667
QY	1382	GCCACTGTGCCCCCAACAGCCCTGGGCGGTACCCAACTCTCCCTCTGCTGCCCTGGGGGC	1441
Db	668	AlaTrpValProHisIGlnAlaLeuGUGUyArgTrpGlnSerProAlaLaProGUGUy	687
QY	1442	ATGAAGACCCGCCAGACCAACCCGCTAACACACCTCTGTACSCCAAGTGGTGTACAGC	1501
Db	688	MetIysSerProProAspRGInProValIySHisLeuPheThrThIGUyAlaValIyTrp	707
QY	1502	ACGTTCATGCTTAAAGACCAACAGTCAGTGGGGAGACACACAGCTGCACCTGAGCATGCT	1561
Db	708	ThrPheMetLeuIySHisIGInCysMetCysGUGUyAsnThrHisValHisProGlnHisAla	727
QY	1562	GCCCGGATCCAGAGCATCTGTGCTCCGGCTGCAGGACACAGGCTGCTTAAGCAATGGCAG	1621
Db	728	GIyArgIleGlnSerIleTrpSerArgLeuGUGUGUyThIGUyLeuSerIyCysGln	747
QY	1622	CGGATCCGAGGTGTCCAAAGCCAGCGTGTGATGAGATGCACAGAGTCGATCAATACAC	1681
Db	748	ArgIleArgGUGUyArgIySalatThrLeuAspGUGUyIleGlnThrValHisSerGUTyHis	767
QY	1682	ACCGTCTCTATGGGAGACCAAGTCCCTTCACCGGACAGAGCTGACAGCAAGATGTCTC	1741
Db	768	ThrLeuLeuIyGUGUyTrpSerProLeuAsnArgGUGUyLeuAspSerIySHisLeu	787
QY	1742	GSTCCCATCCACCAGAGAGTGTATGCTGTCTCTCCCTTGTTGGGGCATGGGGGTGACAGT	1801
Db	788	GIyProIleSerGInIySHisMetIyAlaValAlaLeuProCysGUGUyIleGUGUyAlaAspSer	807
QY	1802	GACACCGTGTGATGTAGATGAGCATCCTCCAGTGGTGTGCGATGGAGTGGGCGTGGCTG	1861
Db	808	AspThrValIyTrpAsnGInMetHisSerSerAlaValAlaIyMetAlaValIyCysLeu	827
QY	1862	CTGAGCTGAGCTTCAAGGTGGCTGCAGGAGAGCTCAAGAAATGGATTTGCCATCATCCGG	1921
Db	828	LeuGInLeuAlaPheIyValAlaAlaIGUGUGUyLeuIyAsnGUGUyPheAlaIleLeuArg	847
QY	1922	CCCCGAGACACCAACGCCGAGGAATCCACAGCCATGGATTTGCTTCTCACTGTGA	1981
Db	848	ProProGUGUyHisAlaIGUGUGUySerThrAlaMetGUGUyPheCysPheHisSerVal	867
QY	1982	GCCATCACCCGAAACCTCCACAGAGGAAGTTGAACGTGGGACAGCTCTCATGCTGGAC	2041
Db	868	AlaIleThrAlaIySHisLeuGInGUGUGUyIySHisLeuAsnValGUGUyValLeuIleValAsp	887
QY	2042	TGGGACATTCACATGTGCAATGGCACCCAGCAGAGCTTGTAAATAGACCCCTGTGTC	2101
Db	888	TrpAspIleHisSHisGUGUyAsnGUGUyThIGInGInAlaPheTyrAsnAspProSerValLeu	907
QY	2102	TACATCTCTGCAATGCGTATGACAAAGGAACTTCTTCCAGGCTGTGGGGCTTCGAA	2161
Db	908	TyrIleSerLeuHisAlaIGUyTrpAsnGUGUyAsnPheProGUGUySerGUGUyAlaProGUGUy	927
QY	2162	GAGGTGTGTGAGGACCAAGCGCGGGGTCAATGTCAACGTGGCATGTGACAGAGAGTGTG	2221
Db	928	GIUValIGUGUGUGUyProGUGUyValIGUyTyrAsnValAsnValAlaIleTrpThIGUGUGUyAla	947
QY	2222	GACCCCCCATTTGAGAGCTGGAGATACCTTACAGCCCTTGAGAGACAGTGTATGGCCATT	2281
Db	948	AspProProIleGUGUyAspValIGUGUyTrpLeuThrAlaPheArgTrpValValIyMetProIle	967
QY	2282	GCCCCAGAGTTTCACTCATGTGTGGTCTAGTCTCCGGCGGGTTGATGTGTGAAGGA	2341

Db	968	AlAHnISgluPhSerProCoSPValValLeuValSerAlaIGlyPheAspAlaValGIuGly	987
Qy	2342	CATGTGTCCCTCTGGTGGTACACTGTGTACACCGCAGATGTTTTGGCCACTTGACACAG	2401
Db	988	HISLeuSerProLeuGlyIleTyrSerValTHrAlaArgCyPheGlyHisLeuTHrArg	1007
Qy	2402	CAGCTGATGACCCTGGCAGAGGGGCCGGTGTGTCTGGCCCTGGAGGAGGCCATGACTTG	2461
Db	1008	GIuLeuMetThrLeuAlaIGlyArgValValLeuAlaLeuGluGlyGlyHisAspLeu	1027
Qy	2462	ACCGCCATGTGTGATGCCCTCTGAAGTGTGTGTCTGGCTCTGTCAAGTGTAAAGCTGCAG	2521
Db	1028	THrAlaIleCyAspAlaSerGluAlaCyValSerAlaLeuLeuSerValGIuLeuGln	1047
Qy	2522	CCCTTGATGAGGACGACTTCTTGACAGCAAAAGCCCAACATCATCAGCAGCATGGCCAGTAAG	2581
Db	1048	ProLeuAspGluAlaValLeuGlnGlnIlyProHisIleAsnAlaValAlaThrLeuGln	1067
Qy	2582	AAAGTCATCGAGATCCAGACCAACACTGTGAGCTGTGTCCAGAAAGTTCGCCCTGTGTG	2641
Db	1068	LysValIleGluIleGlnSerLysHisTrpSerCyValGIuIlyPheAlaIleGlyLeu	1087
Qy	2642	GGCCGGTCCCTCGAGAGGGCCCAAGCAGGTGAGACCCGAGAAAGCCCAAT -GTGAAGCC	2700
Db	1088	GIlyArgSerLeuArgGluAlaGlnAlaGlyGluThrGluAlaGluThrValSerAla	1107
Qy	2701	ATGGCCCTTGCTGTGTGGGGCCGCAAGCCAGCCCAAGCTGTGGCAGCCCGGGAACACAGC	2760
Db	1108	MetAlaLeuLeuSerValGIlyAlaGluGlnAlaGlnAlaAlaAlaAlaAlaArgGluHisSer	1127
Qy	2761	CCCAGCGCGGACAGAGAGCCCATGTGAGCAGAGACCTGGCCCTG 2802	
Db	1128	ProArgProAlaGluGluProMetGluGlnGluProAlaLeu 1141	
RESULT 2			
ID	AAy07092	standard; Protein; 897 AA.	
XX	AC	AAy07092;	
XX	DT	02-JUL-1999 (first entry)	
DE	XX	Colon cancer associated antigen precursor sequence.	
XX	KM	Cancer associated antigen; diagnosis; research; treatment; human;	
KM	XX	breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;	
KM	XX	prostate cancer.	
OS	XX	Homo sapiens.	
XX	PN	WO9004265-A2.	
XX	PD	28-JAN-1999.	
XX	PF	15-JUL-1998; 98WO-US14679.	
XX	PR	22-JUN-1998; 98US-0102322.	
XX	PR	17-JUL-1997; 97US-0896164.	
PR	XX	10-OCT-1997; 97US-0061599.	
PR	XX	10-OCT-1997; 97US-0061765.	
PR	XX	11-OCT-1997; 97US-0948705.	
XX	XX	11-OCT-1997; 97GB-0021697.	
PA	XX	(LUDM-) LUDWIG INST CANCER RES.	
XX	PI	Chen Y, Gout I, Gure A, O'Hare M, Opatka Y, Old LJ;	
PI	XX	Pfeundschuh M, Sahin U, Scanlan MD, Stockert E;	
PI	XX	Tureci O;	
DR	XX	WPI: 1999-132448/11.	
XX	XX	New isolated cancer associated nucleic acids and polypeptides -	

PI Isolated using sera from cancer patients, used to develop products
PI for the diagnosis, monitoring or treatment of cancers
XX
PS Disclosure: Page 652-654; 787pp; English.

CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

SQ Sequence 897 AA;

Alignment Scores:
Pred. No.: 1.66e-306 Length: 897
Score: 4669.00 Matches: 896
Percent Similarity: 99.89% Conservative: 0
Best Local Similarity: 99.89% Mismatches: 1
Query Match: 87.34% Indels: 0
DB: Gaps: 0

US-09-502-945-2 (1-2885) x AAY07092 (1-897)

QY 2 GAATTCCTTTGTGCAAGTCAAGAGAGCCACACAGGCGGCTCAACATTCCCTCCA 61
DB 1 GIuphLeuLeuSerLysSerLysGluProThrProGlyGlyLeuAsnHisSerLeuPro 20
QY 62 CAGCACCCCAATGCTGGGAGCCACCATGCTCTTTGGACAGATTCCCTCCAG 121
DB 21 GlnHisProLysCysTyrPoliValHisHisAlaSerLeuAspGlnSerProGln 40
QY 122 AGCGGCCCCCTGGGAGCGCTCCCTCTCAAAAGCTGTCGCGGGCCCTACAGCACT 181
DB 41 SerGlyProProGlyThrProProSerTyrLysLeuProLeuProGlyPro**AspSer 60
QY 182 CGAGAGCACTCCCTCCGCAAAACAGCCTCTGAACCCCACTTGAATGGCTTCAAG 241
DB 61 ArgAspAspPheProLeuArgLysThrAlaSerGluProAsnLeuLysValArgSerArg 80
QY 242 CTAAACAGAGAGTGGCTGAGCGAGAGACAGTCCCTCTCGCTGCGAGATGGGACT 301
DB 81 LeuLysGlnLysValAlaGluArgArgSerProLeuLeuArgArgLysAspGlyThr 100
QY 302 GTTATTAGACCTTTAAGAGAGAGCTGTGATCACAGGCGCGGCTGGGGGCTG 361
DB 101 ValHisSerThrPheLysArgAlaValAlaGluIleThrGlyAlaGlyProGlyAlaSer 120
QY 362 TCCGCTGTGAACAGGAGCCGCGCTCCGCGCCAGCTCTCCCAAGAGCTCCCAAGCAC 421
DB 121 SerValCysAsnSerAlaProGlySerGlyProSerSerProAsnSerHisSerThr 140
QY 422 ATCGCTGAGATGGCTTTACTGCTCAGTCCCAATCCCACTGAGATGCTCCCTAG 481
DB 141 IleAlaGluAsnGlyPheThrGlySerValProAsnIleProThrGluMetLeuProGln 160
QY 482 CACGAGGCGCTCCCTGGAGAGCTCCCAACAGCTTACAGCTTACAGCTCCCTCT 541
DB 161 HisAlaGluAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTyrThrSerProSer 180
QY 542 CTGCCCAATCTCCCTAGAGGCTGAGGCGACAGCTACTGTCACCAACTCACACCTCACT 601
DB 181 LeuProAsnIleSerLeuGlnLysGlnAlaThrValThrValThrAsnSerHisLeuThr 200
QY 602 GCCTCCCGCAAGCTGTGCACACAGCAGAGGCGGAGAGGCGCTCCACTGCTCGCG 661

DB 201 AlaSerProLysLeuSerThrGlnGlnGlnAlaValArgGlnAlaLeuGlnSerLeuArg 220
QY 662 CAGGCTGGACCCCTACCGGCAAGTTCATGACACATCCTCATTCCTGGCTGGCTG 721
DB 221 GlnGlyGlyThrLeuThrThrLysPheMetSerThrSerSerLeuProLysLeuLeu 240
QY 722 GCGCTGGCACTGAGAGGCGAGGAGCCCAAGGCGCATGCTCCCTGCGAGCATGTG 781
DB 241 GlyValAlaLeuGlnGlyAspLysSerProHisGlyHisAlaSerLeuLeuHisVal 260
QY 782 CTGTTGCTGGACAGGCGCGGACAGACAGACCCCTCATTCCTGTCGACCTCCAGGCGAG 841
DB 261 LeuLeuLeuGlnGlnAlaValArgGlnGlnSerThrLeuIleAlaValProLeuHisGlyGln 280
QY 842 TCCCACTAGTACAGGGGTGAAGCTGTGGCCACACATGCGGAGGTGAGGAGCTCCCG 901
DB 281 SerProLeuValThrGlyValArgValAlaThrSerMetArgThrValGlyLysLeuPro 300
QY 902 CGGCAATCGGCCCTGAGCGCCACTAGTCCCTACCGCTGCGCCGACAGTCCCGGCGCTG 961
DB 301 ArgHisArgProLeuSerThrArgThrGlnSerSerProLeuProGlnSerProGlnAlaLeu 320
QY 962 CAGCAGCTGCTCATGCAACACAGCAGCAGAGTTCTTGAGAGACAGACAGCAGCAG 1021
DB 321 GlnGlnLeuValMetGlnGlnGlnGlnHisGlnPheLeuGlnLysGlnGlnGlnGln 340
QY 1022 CTACAGCTGGGCAAGATCCTCAACAAAGACAGGAGGCTGCCAGGACGCGCCACCCAC 1081
DB 341 LeuGlnLeuGlyLysLysLeuThrLysThrGlyGlyLeuLeuProArgGlnProThrThrHis 360
QY 1082 CCTGAGAGACAGAGAGAGAGCTGACGAGCAGCAGAGAGCTTCTGCGGAGGAGGCC 1141
DB 361 ProGlnGlnThrGlnGlnGlnLeuThrGlnGlnGlnGlnGlnValLeuLeuGlyLysAla 380
QY 1142 CTGACCATGCCCCGGGAGGGGCTCCACAGAGATGAGAGACACACAGAAAGCTGAGAG 1201
DB 381 LeuThrMetProArgGluLysLysSerThrGlnSerGlnSerThrGlnGlnLysLeuGlnGln 400
QY 1202 GAGCAGAGAGAGAGATGGGAGAGAGAGAGAGATGGCATCCAGTTAAGACAGCAGAG 1261
DB 401 GluAspGlnGlnGlnLysAspLysGlnGlnGlnGlnLysAspCysLysGlnValLysAspGlnGln 420
QY 1262 GCGCAGAGTGTGCTGAGAGGCGCCGACTTGAAGAGACCTGCTGGATCAAAAAA 1321
DB 421 GlyGlnSerGlyAlaGlnGlnGlnGlnProAspLeuGlnGlnProGlyAlaGlyLysLys 440
QY 1322 CTGTTCTCAGATGGCCCAACCGCTGCAACCTTGCAAGTGTACCAAGCGCCCTCAGCGCTG 1381
DB 441 LeuPheSerAspAlaGlnProLeuGlnProLeuGlnValTyrGlnAlaProLeuSerLeu 460
QY 1382 GCCACTGTCCCCACCAAGCCCTGGGCGCTTACCAATCCTCCCTGCTGCCCTGGGGCG 1441
DB 461 AlaThrValProHisGlnAlaLeuGlnLysArgThrGlnSerSerProAlaAlaProGlyGly 480
QY 1442 ATGAAGAACCCCAACAGCAACCCGCTCAAGCACTCTTCCACCAAGTGTGTTACGAC 1501
DB 481 MetLysAsnProProAspPleProValLysHisLeuPheThrThrSerValValTyrAsp 500
QY 1502 ACGTTCTATGCTAAACACACAGTGCATGTGCGGGAACACACAGCTGACACCTGAGATGCT 1561
DB 501 ThrPheMetLeuLysHisGlnLysMetCysGlyAsnThrHisValHisProGlnHisAla 520
QY 1562 GCGCGATCCAGAGATGCTGTCGCGGCTGCGAGAGACAGGCTGCTTACGAAGTGCAG 1621
DB 521 GlyArgIleGlnSerThrLeuThrArgLeuGlnGlnThrGlyLeuLeuSerLysCysGln 540
QY 1622 CGGATCCAGAGTGCAGAAAGCCAGCTAGATGAGATCCAGACAGTCCACTGTGAATACAC 1681
DB 541 ArgIleArgGlyArgGlyAlaThrLeuAspGlnIleGlnThrValHisSerGlyLysHis 560
QY 1682 ACCCTGCTCTATGGAGCAGTCCCTCAACCGGCGAAGCTTAGACAGCAAGAGTTGCTC 1741

Db 561 ThrLeuLeuYrGIyThrSerProLeuAnArgInLysLeuAspSerLysLeuLeu 580
 QY 1742 GGTCCATCACCAGAGATGATGCTGTGCTGCTGTGGGGCATGGGGTGACAGT 1801
 Db 581 GILProLeuSerGlnLysMetTryAlaValLeuProCysGlyLysIleGlyValAspSer 600
 QY 1802 GACACCGTGTGGAATGAGATGCACCTCTCCAGTGTGTGCGCATGGCAGTGGCTGTG 1861
 Db 601 AspThrValTrpAsnGlnMetHisSerSerSerAlaValArgMetAlaValGlyCysLeu 620
 QY 1862 CTGAGAGCTGGGCTTCAGAGTGGCTGCAGAGAGACTCAAGATGGATTGGCATATCCGG 1921
 Db 621 LeuGlnLeuAlaPheLysValAlaAlaGlyLysLeuLysAsnGlyPheAlaIleLeuArg 640
 QY 1922 CCCCAGAGACACACGGCGAGAGAAATCCACAGCATGGAGATTCTGCTTCACTGTA 1981
 Db 641 ProProGlnHisHisAlaGlnGlnSerThrAlaMetGlyPheCysPhePheAsnSerVal 660
 QY 1982 GCCATCACCGCAAAATCTCTACAGACAGAAGTTGAACGTGGGCAAGTCTCATGTGAC 2041
 Db 661 AlaIleThrAlaLysLeuLeuGlnGlnLysLeuAsnValGlyLysValLeuIleValAsp 680
 QY 2042 TGGAGCATTCACCATGGCAATGSCACCCAGAGCGCTTCAATGACCCCTGTGTCTC 2101
 Db 681 TrpAspIleHisHisGlyAsnGlyThrGlnGlnAlaPheTyraAsnAspProSerValLeu 700
 QY 2102 TACATCTCTGCAATCGGTATGACAGCGGAATCTTCCAGGCTGTGGGGCTCTGAA 2161
 Db 701 TyrIleSerLeuHisAlaGlyTrpAspAsnGlyAsnPhePheProGlySerGlyAlaProGln 720
 QY 2162 GAGGTGTGTGAGAGACACGCGTGGGTGACATGTGAACSTGGATGGAGACAGAGTGTG 2221
 Db 721 GlnValGlyGlyProGlyValGlyTrpAsnValAsnValAlaIleTrpHisGlyVal 740
 QY 2222 GACCCCCCATGGAGACSTGAGTACCTTACACCTTCAGAGACAGTGGTATGCCATT 2281
 Db 741 AspProPheIleGlyAspValGlyIleTyrLeuThrAlaPheArgThrValIleMetProIle 760
 QY 2282 GCCACAGAGTCTGACSTGATGATGGTGTCTACTCCCGGGGTTGATGCTGTGAAGA 2341
 Db 761 AlaHisGlnPheSerProAspValValLeuValSerAlaGlyPheAspAlaValGlnGly 780
 QY 2342 CATCTGTCTCTCTGGGTGCTACTCTGTCTACCGCCAGATGTTTGGCCACTGTGACAG 2401
 Db 781 HisLeuSerProLeuGlyGlyTyrSerValThrAlaArgCysPheGlnHisLeuThrArg 800
 QY 2402 CAGCTGATGACCTGSCAGAGGGGCGGGTGTGCTGGCCCTGAGAGGACCATGACTTG 2461
 Db 801 GlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGlnGlyGlnHisAspLeu 820
 QY 2462 ACCGCGCATGCTGATGCTGTGAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2521
 Db 821 ThrAlaIleCysAspAlaSerGlnAlaCysValSerAlaLeuLeuSerValLysLeuGln 840
 QY 2522 CCGTGTGATGAGGAGTCTTGCACAAAGCCCAACATCAAGCAGTGGCCACCTGAGAG 2581
 Db 841 ProLeuAspGlnAlaValLeuGlnGlnLysProAsnIleAsnAlaValAlaThrLeuGln 860
 QY 2582 AAAGTCATCGAGATCCAGAGCAAACTGAGCTGTGTGCGAAGTTGCCCGCTGTGTG 2641
 Db 861 LysValIleGlnIleGlnIleGlnSerLysHisTrpSerCysValGlnLysPheAlaAlaGlyLeu 880
 QY 2642 GCGCGGTCCTCGAGGGGCGCCAGAGGTGAGACCGCAAGAGCGCAAAATG 2692
 Db 881 GlyArgSerLeuArgGlyAlaGlnAlaGlyLysIleThrGlnGlnAlaGlnMet 897
 RESULT 3
 AAB49958 standard; Protein; 716 AA.
 AC AAB49958;
 XX 08-MAR-2001 (first entry)

XX XX Human histone deacetylase HDAC-5.
 DE
 XX
 KW Histone deacetylase; HDAC-1; HDAC-2; HDAC-3; HDAC-4; HDAC-5; HDAC-C;
 KW HDAC-D; cell cycle; tumorigenesis; cancer; inhibitor; antisense;
 KW gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN M0200071703-A2.
 XX
 PD 30-NOV-2000.
 XX
 PF 03-MAY-2000; 2000MO-IB01252.
 XX
 PR 03-MAY-1999; 99US-0132287.
 XX
 PA (METH-) METHYLENE INC.
 PI Macleod AR, Li Z, Beerman JM;
 XX
 DR WPI: 2001-016407/02.
 DR N-PSDB; AAC89558.
 XX
 PT Antisense oligonucleotide that inhibits expression of a histone
 PT deacetylase, useful for treating and/or alleviating the symptoms of
 PT neoplasia, or for inhibiting neoplastic cell growth in an animal -
 XX
 PS Disclosure; Page 62-64; 125pp; English.
 XX
 CC The present invention provides inhibitors of histone deacetylase enzymes
 CC such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
 CC inhibitors may be antisense strands or they may be compounds identified
 CC by contacting the enzyme with the compound and measuring the resulting
 CC enzyme activity. These inhibitors are useful for treating cancers and for
 CC identifying which histone deacetylase is involved in a neoplasia.
 XX
 SQ Sequence 716 AA;
 Alignment Scores:
 Pred. No.: 2,03e-238 Length: 716
 Score: 3659.00 Matches: 709
 Percent Similarity: 99.44% Conservative: 3
 Best Local Similarity: 99.02% Mismatches: 4
 Query Match: 68.44% Indels: 1
 DB: 22 Gaps: 0
 US-09-502-945-2 (1-2885) x AAB49958 (1-716)
 QY 656 CTGGGCGAGGGTGGCAGCGTGCAGCGCAAGTTCATGAGCACATCTATTCCTGGCTGC 715
 Db 1 LeuArgGlnGlyGlyThrLeuThrGlyLysPheMetSerThrSerIleProGlyCys 720
 QY 716 CTGCTGGCGGTGACCTGAGAGGGGAGCGGAGCCCGCCAGCGGCAATGCTCTGCTGCAG 775
 Db 21 LeuLeuGlyValAlaLeuGlnGlyAspGlySerProHisGlnHisAlaSerLeuLeuGln 780
 QY 776 CATGTGCTGTGTGCTGAGACAGAGCGCGGCGAGAGCAACCCATGCTGCTGTGCTGCAC 835
 Db 41 HisValLeuLeuLeuGlnGlnAlaArgGlnGlnSerThrLeuIleAlaValProLeuHis 840
 QY 836 GGGCAGTCCCGCATGATGAGCGGTGAACGTGTGCGCCACAGCATGGCGAGGTGAGCAAG 895
 Db 61 GlyIleSerProLeuValThrGlyGlnArgValAlaThrSerMetAlaThrValGlyLys 900
 QY 896 CTCCCGGCGCATCGCCCTGAGCGGCACTGCTCTCACCGCTGCCAGAGTCCCGCAG 955
 Db 81 LeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnInsProGln 100
 QY 956 GCGCTGCAGACAGCTGTGTCATGCAACACAGCAGCAGCAGTTCCTGTGGAAGCAGAACAG 1015
 Db 101 AlaLeuGlnGlnLeuValMetGlnGlnGlnHisGlnGlnPheLeuGlnLysGlnLysGln 120

QY 1016 CAGCAGCTACAGCTGGGCAAGATCTCACAGACAGGGGAGCTGCCAGGAGCCACC 1075
 |||||
 Db 121 GlnGlnLeuGlnLeuGlnLysIleLeuThrLysThrGlyIuLeuProArgGlnProThr 140
 QY 1076 ACCGACCTGAGAGACAGAGAGAGAGCTGACGAGACAGAGAGCTTCTGGGGGAG 1135
 |||||
 Db 141 ThrHisProGlnGlnThrGlnGlnGlnLeuThrGlnGlnGlnIleValLeuGlnGly 160
 QY 1136 GAGACCTGACATGCCCCGGAGGAGCTCCACAGAGAGTGAAGACACAGAGAGCTG 1195
 |||||
 Db 161 GlyAlaLeuThrMetProArgGlnGlnGlnSerThrGlnSerIuSerThrGlnGlnAlaPhe 180
 QY 1196 GAGAGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTCACAGTTAAAGAC 1255
 |||||
 Db 181 GlnGlnGlnAlaSerGlnGlnGlnAlaSerGlnGlnGlnGlnGlnGlnGlnGlnGln 200
 QY 1256 GAGAGAGGAGAGAGTGGTGGTGAAGAGGAGGAGGAGTGGAGAGAGAGCTGGTGGATAC 1315
 |||||
 Db 201 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 220
 QY 1316 AAAAACTGTCTCAGATGCCCCAACGCTGCAACCTTTGAGGTGTACAAAGCCCTTC 1375
 |||||
 Db 221 LysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnValGlnAlaProLeu 240
 QY 1376 AGCTGGCCACTGTGCCCCAACGCTGGGCGGTACCAATTCCTCCCTGCTGCCCT 1435
 |||||
 Db 241 SerLeuAlaThrValIProHisGlnAlaLeuGlyArgThrGlnSerProAlaAlaPro 260
 QY 1436 GGGGAGATGAAGAACCCCGACAGAACCCGTCAGACACTTTCACACAGAGTGGTGC 1495
 |||||
 Db 261 GlyGlyMetLysSerProProArgGlnProValLysHisLeuPheThrThrGlyValAl 280
 QY 1496 TACGACACCTTCAATGCTAAAGACACAGTGCATGCGGGGAACACACAGCTGACCTGAG 1555
 |||||
 Db 281 TyrAspThrPheMetLeuLysHisGlnCysMetCysGlyAsnThrHisValHisProGln 300
 QY 1556 CATTGGGCGCGGATCCAGAGCAATCTGTCGCCGGCTCAGAGACAGAGCCCTGTACAG 1615
 |||||
 Db 301 HisAlaGlyArgIleGlnSerIleThrSerArgLeuGlnIleThrIleLeuSerLys 320
 QY 1616 TGGCAGCGATCCGAGGTGSCAAAGCCAGCTAGATGAGATCCAGACAGTCACTGAA 1675
 |||||
 Db 321 CysGlnArgIleArgIleArgLysAlaThrLeuAspGlnIleGlnThrValHisSerGln 340
 QY 1676 TACACACCTGCTATGAGGAGCAAGTCCCTCAACCGGACAGAGTGAACACAGCAAG 1735
 |||||
 Db 341 TyrHisThrLeuLeuThrGlyThrSerProLeuAsnArgGlnIleLysLeuAspSerLys 360
 QY 1736 TTGCTGGGTCCCATCAAGCAAGATGTATGCTGCTGCTGCTGGGGCATCGGGGTG 1795
 |||||
 Db 361 LeuLeuGlyProIleSerGlnLysMetIleValValLeuProCysGlyIleGlyVal 380
 QY 1796 GACAGTGAACCGTGTGAATGATGACACTCTCCAGTGCCTGGGCAATGGCAGTGGGG 1855
 |||||
 Db 381 AspSerAspThrValIleTrpAsnGlnMetHisSerSerSerAlaValAlaArgMetAlaValGly 400
 QY 1856 TGCCTGCTGAGAGTGGCTTCAAGTGGCTGACAGAGAGCTCAAGATGGATTTGCCATC 1915
 |||||
 Db 401 CysLeuLeuGlnLeuAlaPheLysValAlaIleGlnGlnLysAsnGlnPheAlaIle 420
 QY 1916 ATCCGGCCCCCAGAGACACAGCCGAGAGATCCACAGCCATGGGATTCGCTTTCAC 1975
 |||||
 Db 421 IleArgProProGlnHisHisAlaGlnGlnGlnSerThrAlaMetIlePheCysPhePheAsn 440
 QY 1976 TCTGTGTCATGACCGCAAAACGCTACAGAGAGATTAAGTGGGCAAGGCTCCATC 2035
 |||||
 Db 441 SerValAlaIleThrAlaLysLeuGlnGlnLysLeuAsnValGlyLysValLeuIle 460
 QY 2036 GTGAGACTGGAGATTCACCATGAGCATGACCCACAGGAGCTTACATGACCCCTCT 2095
 |||||
 Db 461 ValAspTrpAspIleHisHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 480
 QY 2096 GTGCTCTACATCTCTGTGATGCTATGACAAAGGAGAACTTCTTTCAGAGCTCTGGGGCT 2155

Db 481 ValLeuThrIleSerLeuHisArgTyrAspAsnGlyAsnPhePheProGlnSerGlyAla 500
 QY 2156 CCGTAGAGCTGTGGTGGAGACAGGCGGGGTACAAATGTGAACCGCATGAGACAGAG 2215
 |||||
 Db 501 ProGlnGlnValGlyGlyProGlnGlyValGlyTyrAsnValAsnValAlaIleTrpThrGly 520
 QY 2216 GGTGTGAGCCCCCATTTGAGAGAGTGAAGTACCTTACAGCTTCAAGAGAGTGTATG 2275
 |||||
 Db 521 GlyValAspProProIleGlyAspValGlnTyrLeuThrAlaPheArgThrValMet 540
 QY 2276 CCGATTGCCCCAGAGTTTTCACCTGATGTGGTCTGTACTGCCCGGGGTTTATGCTGT 2335
 |||||
 Db 541 ProIleAlaHisGlnPheSerProAspValValLeuValSerAlaGlyPheAspAlaVal 560
 QY 2336 GAGAGACATGCTCTGCTGGGTGAGTACTGTGTACCCGCGAGATGTTTGGCCATTG 2395
 |||||
 Db 561 GlnGlyHisLeuSerProLeuGlnGlyTyrSerValThrAlaArgCysPheGlnHisLeu 580
 QY 2396 ACCAGGACCTGATGACCTGACAGGGGGCGGGTGTGCTGGCCCTGAGAGAGGACCAT 2455
 |||||
 Db 581 ThrArgGlnLeuMetThrLeuAlaGlnGlyArgValValLeuAlaLeuGlnGlnGlnHis 600
 QY 2456 GACTTGAACGCGCATCTGTGTATGCTGTGAAGCTTGTGTCTGCGCTGCTCACTGTAAAG 2515
 |||||
 Db 601 AspLeuThrAlaIleCysAspAlaSerGlnAlaCysValSerAlaLeuLeuSerValGln 620
 QY 2516 CTGACAGCCCTTGGATGAGGACAGTCTTGCAGCAAAAGCCCAACATCAACGACGTGACAG 2575
 |||||
 Db 621 LeuGlnProLeuAspGlnValAlaValLeuGlnGlnGlnLysProAsnIleAsnAlaValAlaThr 640
 QY 2576 CTAGAGAAAGTATCGATGTCAGAGCAAAAGTGAAGTGTGTGTCAGAAATTCGCCCT 2635
 |||||
 Db 641 LeuGlnLysValIleGlnIleGlnSerLysHisIleThrSerCysValIleGlnLysPheAlaIle 660
 QY 2636 GGTCTGGGCGCGTCCCTGGGAGGAGGCGGCAAGAGTGAAGCCGAAAGCGGAAAT-GTG 2694
 |||||
 Db 661 GlyLeuGlyArgSerLeuArgIleAlaGlnAlaGlyIleThrGlnGlnAlaGlnThrVal 680
 QY 2695 AACGCCATGGCTTGTGTTGGTGGGCGGCAAGAGCCCAAGCTCGGACGCCGAGGA 2754
 |||||
 Db 681 SerAlaMetAlaLeuLeuSerValGlyAlaGlnGlnAlaGlnAlaIleAlaIleArgGln 700
 QY 2755 CACAGCCCCAGGCGGACAGAGAGCCCATGAGACAGAGAGCTGCCCTG 2802
 |||||
 Db 701 HisSerProArgProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 716
 RESULT 4
 AA099660
 ID AA099660 standard; Protein; 717 AA.
 AC AA099660;
 XX
 DT 07-OCT-2002 (first entry)
 DE Human Histone deacetylase isoform 5.
 XX
 KW Human; histone deacetylase; HDAC-5; cancer; cytostatic;
 KW antisense; tumour suppressor; cell proliferation; tumour;
 KW programmed cell death; necrotic cell death.
 OS Homo sapiens.
 OS
 FH Key location/Qualifiers
 FH Misc-difference 102..193
 FT /note="Encoded by GAGAGAGAT"
 FT Misc-difference 340..342
 FT /note="Encoded by TACCAAC"
 FT Misc-difference 434
 FT /note="Encoded by ACG"
 FT Misc-difference 551..553
 FT /note="Encoded by GTCTTA"
 FT Misc-difference 628


```

DB 440 AasnSerValAlaIleThraLalaLysLeuGlnGlnLysLeuAsnValGlyLysValLeu 459
QY 2033 ATCGGAGCATGGAGCATTCACCATGGCAATGGACACCCAGGCGCTTCACATGACCCC 2092
DB 460 lIeValAspTrpAspIleHisIshISglYasnGlyThrGlnGlnAlaPheIlyrAsnAspPro 479
QY 2093 TCTGTGCTCTACATCTCTCTGCACTGATGACAAACGGGAGACTTTCTTCAGGCTCTGGG 2152
DB 480 SerValLeuTyrlIeSerLeuHisArgIlyrAspAsnGlyAsnPheProGlySerGly 499
QY 2153 GCTCCTGAAGAGGTGGTGGAGAGCAGCGCTGGGGTAAATGTCAGCTGGCATGGACA 2212
DB 500 AlaProGlnGlnValGlyGlyProGlyValGlyTyTrAsnValAsnValAlaTrpThr 519
QY 2213 GGAGGTGTGAGCCCCCGCTTGAGAGCTGAGAGTACCTACAGCCCTTGAGAGAGTGGT 2272
DB 520 GlyGlyValAspProProIleGlyAspValGlyTyTrLeuThrAlaPheArgThrValAl 539
QY 2273 ATGCCCATTTGCCAGAGATTCTCACCTGATGTGTC--CTAGTCTCCGCGGGTTGAT 2329
DB 540 MetProIleAlaHisGlyPheSerProAspValAlaThrLeuValSerAlaGlyPheAsp 559
QY 2330 GCTGTGGAAGACATCTGTCTCTCTGGGTGGCTACTCTGTACCGCCGAGATGTTTGGC 2389
DB 560 AlaValAlGlnGlyHisLeuSerProLeuGlnGlyTyTrSerValThrAlaArgCysPheGly 579
QY 2390 CACTTGACACGAGCAGCTGATGACCCCTGGAGGGGCGGCTGGTGGCTGGAGGGA 2449
DB 580 HisLeuThrArgGlnLeuMetThrLeuAlaGlyGlyArgValValLeuAlaLeuGlnGly 599
QY 2450 GGCATGATGATGACCGCCATCTGTGATGCTGTGAAGCTTGCTGTCTGTCTGTCTGTCT 2509
DB 600 GlyHisAspLeuThrAlaIleCysAspAlaSerGlnAlaCysValSerAlaLeuLeuSer 619
QY 2510 CTAAAGCTTCACCCCTTGATGATGAGCAGCTTCTGCAGCAAAACCCCAACTCAACGAGTG 2569
DB 620 ValGlnLeuGlnProLeuAspGlnLeuValLeuGlnGlnLysProAsnIleAsnAlaVal 639
QY 2570 GCCAGCTGAGAAAGTCTATCGAGATCCAGAGCAAAACACTGGAGCTGTGAGAGATGC 2629
DB 640 AlaThrLeuGlnLysValIleGlnThrGlnSerLysHisTrpSerCysValGlnLysPhe 659
QY 2630 GCCGCTGTCTGGCGCGATCCCTGGAGGGGCGCAAGCAGGTGAGACCGCAAGAGCCGAA 2689
DB 660 AlaAlaGlyLeuGlyArgSerLeuArgIuaIaGlnAlaGlyGlnThrGlnGlnAlaGln 679
QY 2690 AT-GTGAACGCCATGGCTTGTCTGTGGTGGGGCGGAACAGGCCCAACTGCGGAGCC 2748
DB 680 ThrValSerAlaMetAlaLeuLeuSerValGlyAlaGlnGlnAlaGlnAlaAlaAla 699
QY 2749 CGGGAACAGAGCCCGAGCGGAGAGAGCCCATGAGAGAGAGAGAGAGAGAGAGAGAG 2802
DB 700 ArgGlnHisSerProArgProAlaGlnGlnPrometGlnGlnPrometGlnGlnPromet 717

```

```

PD 30-NOV-2000.
XX
XX 03-MAY-2000; 2000MO-IB01252.
XX
XX 03-MAY-1999; 99US-0132287.
XX
XX (METH-) METHYLENE INC.
XX
XX Macleod AR, Li Z, Besterman JM;
XX
XX WPI: 2001-016407/02.
XX
XX N-PSDB: AAC89557.
XX
XX Antisense oligonucleotide that inhibits expression of a histone
XX deacetylase, useful for treating and/or alleviating the symptoms of
XX neoplasia, or for inhibiting neoplastic cell growth in an animal -
XX
XX Disclosure: Page 56-59; 125pp; English.
XX
XX
XX The present invention provides inhibitors of histone deacetylase enzymes
XX such as HDAC-1, HDAC-2, HDAC-3, HDAC-4, HDAC-5, HDAC-C and HDAC-D. These
XX inhibitors may be antisense strands or they may be compounds identified
XX by contacting the enzyme with the compound and measuring the resulting
XX enzyme activity. These inhibitors are useful for treating cancers and for
XX identifying which histone deacetylase is involved in a neoplasia.
XX
XX
XX Sequence 967 AA:
XX
XX Alignment Scores:
XX Pred. No.: 4,24e-173 Length: 967
XX Score: 2691.00 Matches: 568
XX Percent Similarity: 71.38% Conservative: 118
XX Best Local Similarity: 59.11% Mismatches: 195
XX Query Match: 50.34% Indels: 81
XX DB: Gaps: 23
XX
XX US-09-502-945-2 (1-2885) x AAB49957 (1-967)
XX
XX QY 2 GAATTCCTCTGTGCAAGTCAAGAGAGCCACAGCGGCGCTCAACCATTCCTCCGCA 61
XX |||||:||||:||||| ||| ||||| |||
XX DB 60 GluPheValLeuAsnLysLysLysAlaLeuAlaHisArgAsnLeuAsnHisCysIleSer 79
XX
XX QY 62 CAGCACCACCAATGCTGG-----GGAGCCACACATGCTTTTGGACCAAGTTCCTCT 115
XX ||||: ||| |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX DB 80 SerAspProArgTyTrPrpIlyrGlyThrGlnHisSerSerLeuAspGlnSerSerPro 99
XX
XX QY 116 CCCCAGAGCGGCCCCCCTCCGAGAGCGCTCCCTCTACAACTGCTTGGCTGGCGGCTAC 175
XX ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 100 ProGlnSer-----GlyValSerThrSerTyTrAsnHisProValLeuGlyMetTyTr 116
XX
XX QY 176 GACAGTCGAGAGACTTCCCTCCGCAAAAGAGCTCTGAACCCCAACTTGAAGTGCST 235
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX DB 117 AspAlaLysAspAspPheProLeuArgLysThrAlaSerGlnProAsnLeuLysLeuArg 136
XX
XX QY 236 TCAAGGCTAAACAGAAAGTGCCTGAGCGAGAGAGAGTCCCTCTGCTGGCAAGAT 295
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX DB 137 SerArgLeuLysGlnLysValAlaGlnArgArgSerSerProLeuLeuArgIlyAsp 156
XX
XX QY 296 GGGACTGTATTATGACACTTTAAAGAGAGAGCTGTGAGATGACAGAGTCCGGGCTGGG 355
XX ||| ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX DB 157 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 171
XX
XX QY 356 GCGTGTCCGCTGTAAACAGCGCACCGGCTCCGGGCCGAGCTTCCC--AACAGCTCC 412
XX ||| ||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX DB 172 ---AspSerAlaCysSerSerThrAlaProGlySerGlyProSerSerProAsnHisSer 190
XX
XX QY 413 CACAGCACCATGCTGAGAGATGCTTTACTGCTCAGTCCCAACACTCCCACTGAGATG 472
XX ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 191 GlySerValSerAlaGlnAsnGlyIleAlaProAlaValProSerIleProAlaGlnThr 210
XX
XX QY 473 CTCCTCAGCAGCAGCGCTCCCTCTGAGACAGCTCCCAACAGTCAAGCTTACAG 532
XX ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
XX DB 211 SerLeuAlaHisArgLeuValAlaArgGlnGlySerAlaAlaProLeuProLeuTyThr 230

```

QY 533 TCFCCTTTGSCCAACATCTCCCTAGAGGCGAGGCGGACAGTCTGACCAACATCA 592
 |||||
 Db 231 SerProSerLeuProAsnIleThrLeuGlyLeuProAlaThr----- 244
 QY 593 CACCTCACTGCTCCCGCAAGCTGTGACA-----CAGCAGAGGCGCGAGAGCGAGCC 646
 |||
 Db 245 -----GlyProSerAlaGlyThrAlaGlyInGlnAspThrGlnAlaLeuThr 260
 QY 647 CTCACAGTCCCTGCGGAG-----GGTGGCAGCTGACCGGCACTTCATG 691
 |||
 Db 261 LeuProAlaLeuGlnGlnAlaArgLeuSerLeuProGlyThrHisLeuThrProTyrLeu 280
 QY 692 AGCAGATCTCATATCTCGGTGCTGCGGCGCGAGCGAGGCGGCGAGGCGGCC 751
 |||||
 Db 281 SerThrSer-----ProLeuGlnAlaGlyAspGlyValAla 291
 QY 752 CACGGGCACTGCTCCCTGCTGACAGATGCTGTGTGTGAGCAGAGCGCGCGAGCAGAGC 811
 |||||
 Db 292 ---AlaHisSerProLeuGlnHisMetValLeuLeuGlnProProAlaGlnAla 310
 QY 812 ACCCTCAT-----GCTGTGCACTCCACGGGCGACTCCCACTAGTGACGGGT 859
 |||||
 Db 311 ProLeuValThrGlyLeuGlyAlaLeuProLeuHisAlaGlnSer---LeuValGlyAla 329
 QY 860 GAAGGTGGCGCAGCAGATCGGAGCGGTAGGCAAGCTCCGCGGAGATCGGCGCTGAGC 919
 |||||
 Db 330 AspArgValSerProSer-----IleHisTyrLeuAlaGlnGlnHisArgProLeuGly 346
 QY 920 CGCAGCTCAGTCTTACCGCTGCGCGAGAGTCCCGAGCGCTCGCAGCAGCTGTCATGCA 979
 |||||
 Db 347 ArgThrGlnSerAlaProLeuProGlnAsnAlaGlnAlaLeuGlnHisLeuValIleGln 366
 QY 980 CAACAGCAGCAGAGTCTCTGAGAGCAGAG-----CAGCAGCAGCTACAG 1027
 |||||
 Db 367 GlnIleHisGlnGlnProLeuGlnTyrHisTyrGlnGlnProGlnGlnGlnGlnGln 386
 QY 1028 CTGGGCAAGATCTTACCAACAGCAGAGGAGCTGCGCGAGCGCGCGCGCGCGCGCTGAG 1087
 |||||
 Db 387 MetAsnTyrIleIleProLysProSerGlnProAlaArgGlnProGlnSerHisProGlu 406
 QY 1088 GAGCAGAGGAGAGCTGACGCGAGCAGCAGAGAGTCTTGTGGGGAGAGAGCCCTG--- 1144
 |||||
 Db 407 GlnThrGlnGlnGlnGlnAlaArgGlnHisGln---AlaLeuLeuAspGlnProTyrLeuAsp 425
 QY 1145 ASCATGCCCCGGAGGGGCTCCACAGAGAGTGAAGAGCAGCAGCAAGCAAGCAGCTGAGAGAG 1204
 |||||
 Db 426 ArgLeuProGlyGlnLysGlnAlaHisAlaGlnAlaGlyValGln---ValLysGlnGln 444
 QY 1205 GACGAGAGAGAGATGGGAGAGAGAGAGATTCATCCAGGTTAAGAGCAGAGAGAGCGC 1264
 |||||
 Db 445 ProIleGlnSerAspGlnGlnGlnAlaGln-----ProProArgGlnVal 459
 QY 1265 GAGAGTGGTGTGAAGAGAGGCGCGACTTGAAGAGCCTGTGTGATCAAAAAGCTG 1324
 |||||
 Db 460 GlnProGlyGlnAlaArgGln---ProSerGlnGlnLeuLeuProLeuArgGlnAlaAla 478
 QY 1325 TTTCAGATGCGCCCAACCGCTTGAAGAGTGAAGAGGAGGCGCGCTGAGCGCTGCGCC 1384
 |||||
 Db 479 LeuLeuGlnGlnGlnAlaGlnIleHisGlnLeuArgAsnTyrGlnAlaSerMetGlnAlaAla 498
 QY 1385 ACTGTGCGC-----CACCAAGCCTTGGCGCTTACCAACCTCCCTGCT 1429
 |||||
 Db 499 GlyIleProValSerPheGlyGlyHisArgProLeuSerArgAlaGlnSerSerProAla 518
 QY 1430 GCC-----CCTGGGCGCATGAGAGACCCCGCAGACCAACCGCTGACAGCCTTCTC 1480
 |||||
 Db 519 SerIleThrPheProIleSerValGlnGlnProProThrLysPro-----ArgPhe 535
 QY 1481 ACCCAAGAGTGGTCTACGACAGCTTCATGCTAAGCAGCAGCAGCTGCGGGAGACCA 1540
 |||||
 Db 536 ThrThrGlyLeuValTyrAspThrLeuMetLeuLysHisGlnCysThrCysGlySerSer 555

QY 1541 CAGCGCACCTGAGCATGTGGCGGAGTCCAGAGCATGTGTCCGCGCTGAGAGACA 1600
 |||||
 Db 556 SerSerHisProGlnHisAlaGlyArgIleGlnSerIleThrPheArgLeuGlnGlnThr 575
 QY 1601 GGCCTGTTCAGCAAGTGGCAGCGGATCGAGAGTGCAGAAAGCCAGATGAGATCCAG 1660
 |||||
 Db 576 GlyLeuArgGlyLysCysGlyCysIleArgGlyThrGlySalThrLeuGlnGlnGln 595
 QY 1661 ACAGTGCACCTGAAATPACACACCTGCTATATGGAGCAGTCCCTCAACCGGACAGAG 1720
 |||||
 Db 596 ThrValHisSerGlnAlaHisThrLeuLeuTyrGlyThrAsnProLeuAsnArgGlnLys 615
 QY 1721 CTAGACAGCAAGAGTGTGCTGCTCCCTGAGCGCCAGAAAGATGATGCTGCTGCTGCT 1780
 |||||
 Db 616 LeuAspSerLysLysLeuLeuGlySerLeuAla---SerValPheValArgLeuProCys 634
 QY 1781 GGGGCGCATCGGGGTGGACAGTGAACACCTGTGAGATGAGATGCACTCCAGTCTGTG 1840
 |||||
 Db 635 GlyGlyValGlyValAlaAspSerAspThrIleThrAsnGlnValHisSerAlaGlyAlaAla 654
 QY 1841 CGCATGCGAGTGGGCTGCTGCTGAGCTGAGCTTCAGAGTGGCTGCAGAGAGACTCAAG 1900
 |||||
 Db 655 ArgLeuAlaValGlyCysValAlaGlyLeuValAlaThrGlyGlyLeuLys 674
 QY 1901 AATGAGATTGCCATCATCGCGCGCCAGAGCAGCAGCGCGAGATCCACAGCCATGGGA 1960
 |||||
 Db 675 AsnGlyPheAlaValAlaArgProProGlyHisHisAlaGlnGlnSerThrProMetCly 694
 QY 1961 TTCTGCTTTTCAACTGTGTAGCCATCAGCGCAAACTCTTACAGCAGAGATGAGCTG 2020
 |||||
 Db 695 PheCysTyrPheAsnSerValAlaValAlaLysLeuLeuGlnArgLeuSerVal 714
 QY 2021 GGCAGGTCCTCATCGTGGAGTGGAGATCAGCATGGCAGATGGCAGCCAGAGCGCTG 2080
 |||||
 Db 715 SerLysIleLeuIleValAspThrAspValHisHisGlyAsnGlyThrGlnIleAlaPhe 734
 QY 2081 TACAATGACCCCTCTGTACTTATCATCTCTGTGATGCTGTGACAGGAGAACTTCTT 2140
 |||||
 Db 735 TyrSerAspProSerValLeuTyrMetSerLeuHisArgTyrAspArgGlyAsnPhePhe 754
 QY 2141 CCAGGCTTGGGGCTCTGAAAGAGTGTGTGTGAGCAGCAGCGCTGGGTACAAATGTGAC 2200
 |||||
 Db 755 ProGlySerGlyAlaProAspGlyValGlyThrGlyProGlyValGlyPheAsnValAsn 774
 QY 2201 GTGCGATGACAGGAGGTGTGAGCCCGCCCATGTGAGAGCAGTGAAGCTTACAGCCCTG 2260
 |||||
 Db 775 MetAlaPheThrGlyGlyLeuAspProProMetGlyAspAlaGlnTyrLeuAlaAlaPhe 794
 QY 2261 AGCAGAGTGTGATGCCATGTGCCAGAGTCTCACCTGATGTGTGCTAGTCTCGCGC 2320
 |||||
 Db 795 ArgThrValAlaMetProIleAlaSerGlnPheAlaProAspValValLeuValSerSer 814
 QY 2321 GGGTTGATGCTGTGAAGAGCATGTGTCTCTGTGGTGGCTACTGTGACCGCCAGA 2380
 |||||
 Db 815 GlyPheAspAlaValGlyGlnHisProThrProLeuGlyGlyTyrAsnLeuSerAlaArg 834
 QY 2381 TGTTTTGCCATCGACACGAGCAGTGTATGACCTGTGGAGAGGCGCGGCTGCTGCGCC 2440
 |||||
 Db 835 CysPheGlyTyrLeuThrLysGlnLeuMetCylLeuAlaGlyGlyAlaLeuAla 854
 QY 2441 CTGAGAGGAGGCGCATGACTTGAACCGCATCTGTATGCTGTGAGCTGTGTCTCGCT 2500
 |||||
 Db 855 LeuGlnGlyGlnHisAspLeuThrAlaIleCysAspAlaSerGlnValLysValSerAla 874
 QY 2501 CTGCTCAGTGAAGAGCTGCGACCGCTTGGATGAGCAGTCTTGCGAGCAAAAGCCCAACATC 2560
 |||||
 Db 875 LeuLeuGlnGlnGlnLeuAspProLeuProGlnLysValLeuGlnGlnArgProAsnAla 894
 QY 2561 AAGCAGTGGCGCAGCTAGAGAAAGTCAATGATGACAGCAACAGCTGAGTGTGCTG 2620
 |||||
 Db 895 AsnAlaValAlaGlySerMetGlnLysValMetGlnIleHisSerLysTyrTrpArgCysLeu 914
 QY 2621 CAGAAATTGCGCGCTGTGCTGCGCGGCTCCCTGAGAGGCGCCCAAGCAGTGAAGCCGAA 2680

```

Db 915 GlnArgThrThrSerThrAlaGlyArgSerLeuIleGluAlaGlnThrCysGluAsnGlu 934
QY 2681 GAAGCCGAAT-GRGAAGCCATGGCCCTTGCTGTGGGGGGCCGGAACAGCCCAAGCT 2739
Db 935 GluAlaGluThrValThrAlaMetAlaSerLeuSerValGlyValLysProAlaGlu--- 953
QY 2740 GCGGACGCCCGGGAACACAGCCCGGCGGAGAGAGAGCCCATGAGAGAGCCCTGCC 2799
Db 954 -----LysArgProAspGluGluProMetGluGluGluProPro 966
QY 2800 CTG 2802
Db 967 Leu 967

RESULT 6
AAU99659 standard; Protein: 967 AA.
ID AAU99659 standard; Protein: 967 AA.
AC AAU99659;
XX
XX 07-OCT-2002 (first entry)
DE Human Histone deacetylase isoform 4.
XX
XX Human; histone deacetylase; HDAC-4; cancer; cytostatic;
KM antisenase; tumour suppressor; cell proliferation; tumour;
XX programmed cell death; necrotic cell death.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 72
FT Misc-difference /note= "Encoded by CCG"
FT Misc-difference 78
FT Misc-difference /note= "Encoded by ACT"
FT Misc-difference 80..81
FT Misc-difference /note= "Encoded by AGAGAC"
FT Misc-difference 135
FT Misc-difference /note= "Encoded by TCA"
FT Misc-difference 194..196
FT Misc-difference /note= "Encoded by GCGTGGAG"
FT Misc-difference 318
FT Misc-difference /note= "Encoded by GTA"
FT Misc-difference 574
FT Misc-difference /note= "Encoded by AAG"
FT Misc-difference 701
FT Misc-difference /note= "Encoded by GCG"
FT Misc-difference 775
FT Misc-difference /note= "Encoded by ACG"
FT Misc-difference 812
FT Misc-difference /note= "Encoded by GTC"
FT Misc-difference 948
FT Misc-difference /note= "Encoded by GAC"
XX
XX US2002061860-A1.
XX
XX 23-MAY-2002.
XX
XX 06-AUG-2001; 2001US-0817913.
XX
XX 24-MAR-2000; 2000US-192157P.
XX
XX (LIZZ/) LI Z.
XX (BONF/) BONFILS C.
XX (BEST/) BESTERMAN J.
XX
XX LI Z, Bonfils C, Besterman J;
XX
XX WPI: 2002-507650/54.
XX DR N-PSDB; ABK87718.
XX
XX Agent that specifically inhibits an isoform of histone deacetylase.

```

```

PT useful for treating cancer and other cell proliferative diseases,
PT preferably comprises an antisense oligonucleotide
XX
XX Disclosure; Fig 4A; 60pp; English.
XX
XX The invention relates to an agent that inhibits an isoform of histone
CC deacetylase (HDAC-1 to HDAC-8) but not all isoforms, e.g. an
CC antisense oligonucleotide. Also included are inhibiting an HDAC isoform
CC in a cell by treatment with the agent, identifying an HDAC isoform that
CC is required for induction of cell proliferation or differentiation and
CC inhibiting cell proliferation by treatment with two antisense
CC oligonucleotides or small molecules that inhibit a specific HDAC
CC isoform, or antisense oligonucleotide or small molecules that inhibit
CC DNA methyltransferase. The agent therefore acts as a tumour suppressor.
CC The agents are used to treat diseases of cell proliferation and
CC differentiation (e.g. cancer and tumours), by inducing growth retardation,
CC growth arrest or programmed/necrotic cell death, specifically neoplastic
CC cell proliferation in humans. The agents are selective for particular
CC isoforms, compared to known inhibitors which are not selective.
CC The present sequence represents the HDAC-4 isoform.
XX
XX Sequence 967 AA;
SQ

```

```

Alignment Scores:
Pred. No.: 1,08e-172 Length: 967
Score: 2685.00 Matches: 567
Percent Similarity: 71.28% Conservative: 118
Best Local Similarity: 59.00% Mismatches: 196
Query Match: 50.22% Indels: 81
Dbs: Gaps: 23

```

US-09-502-945-2 (1-2885) x AAU99659 (1-967)

```

QY 2 GAATCTCTGTGTGAGTCAAGAGAGCCACACAGGCGGCTCAGCATTCCTCCCA 61
Db 60 GlupheValIleuAsnLysLysLysAlaLeuAlaHisProAsnLeuAsnHisCysIleSer 79
QY 62 CAGCACCCCAATGCTGG-----GGAGCCCAAGCTCTTTGGACAGAGTCCCT 115
Db 80 SerCysProArgTyrTrpTyrGlyLysThrGlnHisSerSerLeuAspGlnSerSerPro 99
QY 116 CCCCAGAGCGGCCCCCGGAGCGCTCCCTCCCAAGTGGCTTCGCGGCGCT 175
Db 100 ProGlnSer-----GlyValSerThrSerTyrAsnHisProValLeuGlyMetLys 116
QY 176 GACAGTCGAGAGACTTCCCTCCGCAAAACAGCCTGTAACCCACTTGAAGTGGCT 235
Db 117 AspaIalysAspaSerPheProLeuArgLysThrAlaSerGluProAsnLeuLysLeuArg 136
QY 236 TCAAGCTTAAACAGAGAGTGGCTGAGCGAGAGACAGTCCCTCTCGTCGCAAGAT 295
Db 137 SerArgLeuLysGlnLysValAlaIleuArgArgSerSerProLeuIleuArgLysAsp 156
QY 296 GGGACTGTATTAGACCTTTAAGAGAGAGAGCTGTGAGATCGAGCGGCGGCTGG 355
Db 157 GlyProValValThrAlaLeuLysLysArgProLeuAspValThr----- 171
QY 356 GCGTGTCCGTGTGTAACAGGAGCCGCTCCGCGCCAGCTCTCC---AACAGCTCC 412
Db 172 ---AspSerAlaCysSerSerAlaProGlySerGlyProSerSerProAsnAsnSer 190
QY 413 CACAGCACCATGCTGAGAAATGGCTTACTGGCTCAGTCCCAACATGCCCACTGAGATG 472
Db 191 GlySerValSerAlaGluAsnGlyLeuAlaProAlaValProSerIleProAlaIleuThr 210
QY 473 CTCCTCAGACCGAGCCCTCCCTGTGACAGAGCTCCCAACACATGAGCTCTCAGACG 532
Db 211 SerLeuAlaHisArgLeuValAlaIleuArgGluGlySerAlaIleuProLeuProLeuThr 230
QY 533 TCTCTCTCTGCCCACATCTCCCTAGGCTGAGCGGCGAGCTCAGTGCACCAACTCA 592
Db 231 SerProSerLeuProAsnIleThrLeuGlyLeuProAlaThr----- 244

```

[illegible]

Dd	576	gLyLeuAhrGlyLyScySeLuStuStlLeArGrLyArdLySaLaThrLeuGluGluLeu	In	595
Qy	1661	ACAGTGCACCTCGAATTAACACACACCCCTGCTATATGAGGACAGACCCCTCAACCGGACAG		1720
Dd	596	ThrValHisSerCjUAlaHisThrLeuLeuYrGjLyThrAsnProLeuAsnAArgInLys		615
Qy	1721	CTAGACACGACAGACTGCTGCTGGTCCCATCAAGCCAGAAAGATGATGCTGCTGCTTGT		1780
Dd	616	LeuAspSerLysLyLeuLeuGluSerLeuAla--SerValPheValArdLeuProCys		634
Qy	1781	GGGGGCATCGGGGTGGACAGTACACCCGCTGGATGAGATGACACTCCCTCCAGTCCCTG		1840
Dd	635	GlyGlyValGlyValAspSerAspThrLLeTrpAsnGluValHisSerAlaGlyAla		654
Qy	1841	CGCATGGCAGTGGGTGGCTGCTGGAGATGGCCCTCAAGSGTGGCGTGCAGAGACTCAAG		1900
Dd	655	ArgLeuAlaValGjLyCySValValAlaLLeuLeuValPheLysValAlaThrGjLyGluLeuLys		674
Qy	1901	AATGATTTGGCATATCCGGCCCCAGACACACCGCCGAGAAATCCACAGCCATGGGA		1960
Dd	675	AsnGlyPheAlaValAlaArProProGjLynHisHisAlaGjLyGluSerThrProMetGjLy		694
Qy	1961	TTCTGCTCTTCAACTGTAAGTCATCAACCGAAACCTCAACAGCACAAGTGAAGCTG		2020
Dd	695	PheCysTyTPheAsnSerValAlaValAlaAlaLysLeuLeuGlnGlnArdLeuSerVal		714
Qy	2021	GGCAGTGCCTCATCGTGGACTGGGACGGGACATTCACATGCGCAATGGCACCCACAGCGCTTC		2080
Dd	715	SerLysLLeuLeuLLeuAlaLsPrpAspValHisHisSgLyAsnGjLyThrGlnGlnAlaPhe		734
Qy	2081	TACATGACCCCTGTGTGCTTACATCTCTGTGCACTGCTATGACACGGAACTTCTT		2140
Dd	735	TyrSerAspProSerValLeuLysMetSerLeuHisAlaArgLysAspSgLyAsnPhePhe		754
Qy	2141	CGAGCTCTGGGGGCTCCGAAGAGGTGTGGAGAGACAGCGGTGGGTACAAATGTGAC		2200
Dd	755	ProGjLySerGjLyAlaProAspLeuValGjLyThrGjLyProGjLyValGjLyPheAsnAlaSn		774
Qy	2201	TGGCATGAGACGAGAGGTGTGGACCCCCCATTTGAGACGTGAGATACCTTAACAGCTTC		2260
Dd	775	MetAlaPheThrGjLyGjLyLeuAsnProProMetGjLyAspAlaGjLyLysLeuAlaAlaPhe		794
Qy	2261	AGGACAGTGGTGAATGCCCATTTGCCACGAGATTTCACCTGATGGTCTAGTCCGCC		2320
Dd	795	ArgThrValAlaMetProLLeAlaSerGjLunPheAlaProAspValValLeuAlaSerSer		814
Qy	2321	GGGTTGATGCTGTGAAGAGACATGTGCTCTCTGGGTGGCTACTGTGTACCGCCGGA		2380
Dd	815	GjLyPheAspAlaValAlaGjLyLsPrProThrProLeuGjLyGjLyTyrAsnLeuSerAlaArg		834
Qy	2381	TGTTTTGGCACTTACACGAGGAGTGAATGACCCCTGGAGGGGGGGGTGGTGTGGCC		2440
Dd	835	CysPheGjLyTyLLeuThrLysGlnLeuMetGjLyLeuAlaGjLyGjLyArdLLeValLeuAla		854
Qy	2441	CTGGAAGGAGGCGCATGACTTGAACCGCCCATCTGTGTATGCTGTGAAGCTTGTGTCCGCT		2500
Dd	855	LeuGjLyGjLyHisAspLeuThrAlaLLeCysAspAlaSerGjLyAlaCySValSerAla		874
Qy	2501	CTGCTCACTGTAAACCTGCAGCCCTTGGATGAGCAGTCTTGCACACAAACCCACATC		2560
Dd	875	LeuLeuGjLyAsnGjLyLeuLeuAspProLeuProGjLyLysValLLeuGlnGlnArdProAsnAla		894
Qy	2561	AACGAGTGGCGACGCTTGAAGAAATGATTCAGAGATCCGACCAAAACACTGGAGCTGTGGT		2620
Dd	895	AsnAlaValArdSerMetGjLyLysValMetGjLyLLeHisSerLysTyTPTrArdCysLeu		914
Qy	2621	CAGAGTTCGCGCGTGTCTGTGGCGGTGCCCTGGCGAGGGGGCCCAAGCAGTGAACAGAA		2680
Dd	915	GjLnArdGjThrHisSerThrAlaGjLyArdSerLeuLLeGjLyAlaGjLyInThrCysGjLyAsnGjLy		934
Qy	2681	GAAGCCGAAT-GTGAACGCCATGGCTTGTGCTGTGGGGGGCCGACAGGCCCAAGCT		2739

DB 935 GIUAlaGluThrValIThrAlaMetAlaSerLeuSerValGIyValIysProAlaIu--- 953
QY 2740 GGGGACGCCGGGACACAGCCCCGCGGACAGAGCCCATGGAGAGAGCCTGCC 2799
DB 954 -----LysArgProAspGluGluProMetGluGluGluProPro 966
QY 2800 CTG 2802
DB 967 Leu 967
RESULT 7
AAM93526
ID AAM93526 standard; Protein; 417 AA.
XX
AC AAM93526;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide; SEQ ID NO: 3261.
XX
KM Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PE 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
XX
N-PSDB: AAK94456.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 3261; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesized by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 417 AA:
Alignment Scores:
Pred. No.: 3,62e-135 Length: 417
Score: 2127.00 Matches: 413
Percent Similarity: 99.28% Conservative: 2
Best Local Similarity: 98.80% Mismatches: 2
Query Match: 39.79% Indels: 1
DB: 22 Gaps: 0
US-09-502-945-2 (1-2885) x AAM93526 (1-417)
QY 470 ATGCTCCTCAGACCGAGCCTCCTCTGACAGCTCCCAACCAAGTTTCAGCCTTAC 529

DB 1 MetLeuProGlnHisArgAlaLeuProLeuAspSerSerProAsnGlnPheSerLeuTy 20
QY 530 AGCTGCTCTTCTCTGCCCCAACAATCTCCCTAGGGCTGCAGGCCACGGTCACTGCACCAAC 589
DB 21 ThrSerProSerLeuProAsnIleSerIleuGlnAlaIleThrValIThrValIThrAsn 40
QY 590 TCACACCTACTGCTGCTCCCGAAGCTGTGCACACAGCAGAGGCCGAGAGGCCCTGC 649
DB 41 SerHisLeuThrAlaSerProIlySerLeuSerThrGlnGlnAlaGlnArgGlnAlaLeu 60
QY 650 CAGTCCCTGCGGCGAGGGTGCACAGCTGCACCGGCAAGTTCATGACACATCTCTATTCTCT 709
DB 61 GlnSerLeuArgGlnGlnIyGlnIleuThrGlnIyLysPheMetSerThrSerSerIlePro 80
QY 710 GGCCTGCTGCTGCGGCTGCGCAGTGCAGGCGGAGGCGGAGGCCCGGAGGATGCTCCG 769
DB 81 GlyCysLeuLeuGlnValAlaLeuGlnIyAspGlySerProHisGlnIleAlaSerLeu 100
QY 770 CTGCAGCATGTGCTGTGTGCTGGAGAGGCCCGGACAGAGCAGACCTCATGTGCTGCGCA 829
DB 101 LeuGlnHisValIleuLeuLeuGlnGlnAlaArgGlnGlnSerThrLeuIleAlaValPro 120
QY 830 CTGCACGGGCGAGTCCCGCACTAGTACGGGTGAACGTGTGGCCACAGCATGCGGAGGTA 889
DB 121 LeuHisGlnIyGlnSerProLeuValThrGlnIyAlaArgValAlaThrSerMetArgThrVal 140
QY 890 GGCAGGCTCCCGGCGATGCGGCCCTGAGCCCGGACAGTCCGTCACCGTGGCGGAGAT 949
DB 141 GlyLysLeuProArgHisArgProLeuSerArgThrGlnSerSerProLeuProGlnSer 160
QY 950 CCCAGGCGCTGCAGCAGCTGTGCATGACAACAACAGCAGCAGCAGTCTCTGAGAGAAGCAG 1009
DB 161 ProGlnAlaLeuGlnGlnIleuValIleuGlnGlnIleuGlnIleuGlnIleuGlnIyGln 180
QY 1010 AAGCAGCAGCAGCTACAGCTGGGCAAGATCTCCACCAACAGGGAGGTGCCACAGCAG 1069
DB 181 LysGlnGlnGlnIleuGlnIleuGlnIyLysIleLeuThrLysThrGlnIleuProArgGln 200
QY 1070 CCCAGCAGCCCGCTGAGAGAGACAGAGAGAGTGCAGCAGCAGCAGAGGTCTGTGCTG 1129
DB 201 ProThrThrHisProGlnIleuThrGlnIleuGlnIleuThrGlnIleuValIleuLeu 220
QY 1130 GGGAGGAGAGCCCTGACCATGCCCCGAGGGCTCCACAGAAGTGAAGCAGACAGAA 1189
DB 221 GlyGlnGlnIyAlaLeuThrMetProArgGlnIySerThrGlnSerIleuSerThrGlnIu 240
QY 1190 GACCTGGAGAGAGAGAGACAGAGAAGAGATGGGGAGAGAGAGAGATGTCATCCAGGT 1249
DB 241 AspLeuGlnIleuGlnIleuAspIleuGlnIleuAspGlyGlnIleuGlnIleuAspGlyIleuVal 260
QY 1250 AAGGACGAGAGAGGCGAGAGTGTGCTGAGAGAGGCGCCGACTTGGAGAGGCTGTGCT 1309
DB 261 LysAspGlnIleuGlnIleuSerGlnIleuGlnIleuAspGlyGlnIleuGlnIleuAspGlyIleuVal 280
QY 1310 GGATACAAAACCTGTTCTCAGATGCCAACCGCTGCACACTTTGGAGGTGTACCAAGCG 1369
DB 281 GlyTyLysLysLeuPheSerAspAlaGlnProLeuGlnProLeuGlnValIyGlnAla 300
QY 1370 CCCCTGAGCCTGGCCACTGTGCCCAACGAGCCCTGGGCGGTACCAATCTCCCTGCT 1429
DB 301 ProLeuSerLeuAlaThrValProHisGlnAlaLeuGlnIyArgThrGlnSerSerProAla 320
QY 1430 GCGCCTGGGGGATGAAGAAGCCCGGACAGCAACCGTAAACACACTTTCACCAACAAT 1489
DB 321 AlaProGlnIyGlnIleuLysSerProProAspGlnProValLysHisLeuPheThrThrGly 340
QY 1490 GTGGTCTACGACAGCTTCATGCTAAAGCAGCAGTGCATGCGGGGAACACAGCTGCAC 1549
DB 341 ValValIyAspThrPheMetLeuLysHisGlnIyThrCysGlnIyAsnThrHisValHis 360
QY 1550 CCTGAGCATGCTGGCGGATGCCAGAGCATGTGCTCCGGCTGCAGAGAGCAGCCTGCTT 1609

Db 361 ProGluHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnIuThrGlyLeuLeu 380
 Oy 1610 AGCAAGTGCAGCGGATCCAGGTCCGCAAGCCAGCTAGATGAGATCCAGAGTGCAC 1669
 |||||
 Db 381 SerIleGlyAlaThrGlyArgIleGlnSerIleTrpSerArgLeuLeu 400
 Oy 1670 TCTGATATCCAGACCCCTCTATGGAGCCAGTCCCTCAACGGCAGAAAC 1721
 |||||
 Db 400 rIeuAsnThrThrProCysSerMetGlyProValProSerThrGlyArgSer 417

RESULT 8

AAM78891

ID AAM78891 standard; Protein; 1008 AA.

XX AAM78891;

XX 06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1553.

XX Human: cytokine; cell proliferation; cell differentiation; gene therapy;

XX tissue growth factor; Immunomodulatory; cancer; leukaemia;

XX nervous system disorder; arthritis; inflammation.

XX Homo sapiens.

XX MO200157190-A2.

XX 09-AUG-2001.

XX 05-FEB-2001; 2001MO-US04098.

XX 03-FEB-2000; 2000US-0496914.

XX 27-APR-2000; 2000US-0560875.

XX 20-JUN-2000; 2000US-0598075.

XX 19-JUL-2000; 2000US-0620325.

XX 01-SEP-2000; 2000US-0654936.

XX 15-SEP-2000; 2000US-0663561.

XX 20-OCT-2000; 2000US-0693325.

XX 30-NOV-2000; 2000US-0728422.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,

XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;

XX Xue AJ, Yang Y, Wejrtman T, Goodrich R;

XX WPI: 2001-476283/51.

XX N-PSDB; AAK52024.

XX Nucleic acids encoding polypeptides with cytokine-like activities,

XX useful in diagnosis and gene therapy -

XX Claim 20; Page 3870-3872; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM7823-AAK60302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhbn activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM60020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.

XX Sequence 1008 AA;

Alignment Scores:

Pred. No.: 3,96e-115 Length: 1008
 Score: 1831.50 Matches: 424
 Percent Similarity: 56.03% Conservative: 124
 Best Local Similarity: 43.35% Mismatches: 249
 Query Match: 34.26% Indels: 182
 DB: 22 Gaps: 22

US-09-502-945-2 (1-2885) x AAM78891 (1-1008)

Oy 11 TTGTCGAAGTCAAAAGAGCCACACAGGCGGCTCACCATTCCCTCCACAGACCCC 70
 ||| : : : : : ||| : : : : :
 Db 157 LeuGluArgThrValHisProAsnSerProGlyIleProIArgThrLeuGluProLeu 176
 Oy 71 AAMTGCCTGGGAGGCCACCATGCTCTTTGGACAGAGTCCCTCCACAGGCGGCC 130
 : : : : : ||| : : : : :
 Db 177 GluThrGluGlyAlaThrArgSerMetLeuSerSerPheLeuProProValProSerLeu 196
 Oy 131 CTTGGAGCGGCTCCCTCTCAAAACTGCTTTCCTGGGCTTACGACAGTGCAGAGC 190
 ||| : : : : : ||| : : : : :
 Db 197 ProSerAspProProGluHis----- 203
 Oy 191 TTCCCTCCCTCGCAAAACAGCCTCTGAACCCACTGAAAGTGGCTTCAAGCTAAACAG 250
 ||||| : : : : : ||||| : : : : :
 Db 204 PheProLeuArgIleThrValSerGluProAsnLeuIleuArgIleGlyProIle 222
 Oy 251 AAGTGGCTGAGCGGAGCAACAGATGCTCCCTGCTGCGCAAGGATGGAGTATTAGC 310
 ||| : : : : : ||| : : : : :
 Db 223 LysSerLeuGluArgIleGlyAsnProLeuLeuArgIleGlySerAlaPro-----Pro 240
 Oy 311 ACCTTTAAAGAGAGAGCTGTGATACAGATCAGTCCGCGGCTGCTCCGTTG 370
 : : : : : : : : : : : ||| : : : : :
 Db 241 SerLeuArgArgArgProAlaGluThrLeuGlyAspSerSerProSerSer----- 258
 Oy 371 AACAGCGCACCCGCGCTCCGCCCCAGCTCTCCCAACAGTCCCAACAGCACCATCCTGAG 430
 ||| ||| ||||| ||||| : : : : :
 Db 259 ---SerThrProAlaSerGlyCysSerSerProAsnAspSerGluHis----- 273
 Oy 431 AATGGCTTACGTGCTACGTCAGTCCCAACATCCCACTGAGTGGTCCCTCAGCACCGAGCC 490
 ||| : : : : : ||| : : : : :
 Db 274 -----GlyProAsnProIleLeuGlySerIleAlaLeuGluGlyIleArgLeu 289
 Oy 491 CTCCTCTGAGACAGCTCCCAACAGTTCAGCTTACAGCTCTCTCTGCTCCCAAC 550
 : : : : : : : : : : : ||| : : : : :
 Db 290 ArgLeuGluIleuThrSerValAlaProPheAlaLeuProThrValSerLeuLeuProAla 309
 Oy 551 ATCTCCCTAGGGGTGACGGCCAGGCTCAGTGCACCAACTGACACTGCTCCCGC 610
 ||| : : : : : ||| : : : : :
 Db 310 IleThrLeuGlyLeuProAla----- 316
 Oy 611 AAGCTGTGCACAGCAGGAGCGCCGAGAGGCGCTCCAGTCCCTGCGGCGAGGTTGC 670
 : : : : : : : : : : : ||| : : : : :
 Db 317 -----ProAlaArgAlaAspSerAspArgArgThrHisProThrLeuGlyProArgGly 334
 Oy 671 ACGGTGACGGCAAGTTCATGAGCAGACATCTCTATTCTGGTGGCTGCGGTGAGCA 730
 : : : : : ||| : : : : :
 Db 335 ProIleLeuGlySerProHisThrProLeuPheLeuProHis-----GlyLeuGlu 351
 Oy 731 CTGAGAGCGCAGCGGAGCCCAACGGGATGCTCCCTGCGAGCATGTGTTGCTG 790
 ||| : : : : : ||| : : : : :
 Db 352 ProGluAlaGlyLeuThr-----LeuProSerArgLeuGluProIleLeuLeu 368
 Oy 791 GAGCAGGCGCGGAGCAGCAGACCTCATGCTGTGTGCATCCAGGCGAGTCCCACTA 850
 : : : : : : : : : : : ||| : : : : :
 Db 369 AspProSerGlySerHisAlaProLeuLeuThrValProGlyLeuGlyProLeuProPhe 388
 Oy 851 GTGACGGGTGAAGCTGTGGCGACACAGCATGAGCAGGTAGCAAGTCCCGGCGATCG 910
 : : : : : : : : : : : ||| : : : : :
 Db 389 HisPheAlaIleuSerLeuMetThrThrGluArgLeuSerIleGly-----GlyLeuHisTrp 407
 Oy 911 CCCCTGAGCGGCACTGATCTCAGCGCTGCGCGAGAGTCCCGAGGCC----- 958
 ||||| : : : : : ||||| : : : : :
 Db 408 ProLeuSerArgThrArgSerGluProLeuProProSerAlaThrAlaProProPro 427

QY	959	-----CTGCAGACAGCTGTGATGCAACAAACAGACACAGACTTCTG	1000
Db	428	gLyProMetgInProArgLeuGluInLeu-----LysThrIsvaInvalIle	444
QY	1001	GAGAAAGCAGAAAGCAGACAGACTACAGTGGGCAATGCTCCACCAAGACAGGAGCTG	1060
Db	445	LysArgSer-----AlaLysProSerTiluys	453
QY	1061	CCGAGAGACCCCAACCCACCTCGAGAGACAGACAGAGAGAGCTGACGAGCAGCAGAG	1120
Db	454	ProArg-----	455
QY	1121	GTCTGTGGGGAGAGGAGCCCTGACCATGCCCCGGAGAGCTCCACAGAGATGAGAC	1180
Db	455	-----	455
QY	1181	ACACAGAAAGACCTGAGAGAGAGACAGAAAGAGATGGAGAGAGAGAGAGATGTC	1240
Db	456	LeuArgInIleProSerAlaGluAspLeuInThrAspGly-----	469
QY	1241	ATCCAGGTTAAAGACAGAGAGGCGGAGAGTGT-----GCTGAGAGAGGCGCCGACATTG	1294
Db	470	-----GlyGlyProGlyInValValAspAspIleuInIHis	482
QY	1295	GAGGAGCCTGGTCGTGGATGCAAAAACTGTCTGATATGCGCAACCGCTGCACCTTGG	1354
Db	483	ArgGluLeuGlnIHisGlyGln-----ProGluInAlaGlyIleProAlaLeuIleu	498
QY	1355	CAGGTGTACCAAGCGCC-----	1372
Db	499	Gln-----GlnIHisProGlnInValLeuLeuTrpGluInGlnIAspLeuAlaGlyArgLeu	516
QY	1373	-----CTCAGCCTGGCCACTGTGTGCCCAACCAAGC	1402
Db	517	ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaInGlyIHisArgPro	536
QY	1403	CTGGGCGCTACCCCAATGCTCCGCTGCTGCGGCGGAGCATGAAAGAACCC-----	1453
Db	537	LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerLarProIuProAla	556
QY	1454	-----CCAGACCAACCCGTACAGCACTC---TTACACCA	1486
Db	557	SerGlnAlaAlaArgValLeuSerSerSerGluThrProAlaArgThrLeuProPheThrIHis	576
QY	1487	AGTGGGTCTGCAGACAGTTCATGCTTAAACACACATGGATGGGGAGACACAGACTG	1546
Db	577	GlyLeuIleTyAspSerValMetLeuLysIHisGlnCysSerGlyIAspAsnSerArg	596
QY	1547	CACCTGTGACATGTGTGCGCCGATGCTCAAGACATGTGTCGCGGCTGAGAGACAGACCTG	1606
Db	597	HisProGlnIHisAlaGlyArgIleGlnIserIleThrSerArgLeuGlnInIAspGlyLeu	616
QY	1607	CTTACAGATGCGAGCGGATGCGAGGTGCGAAAGCCAGCATGATAGATCCAGACATG	1666
Db	617	ArgSerGlnCysGlnCysLeuArgGlyArgGlyAlaSerLeuGlnCysLeuInIserVal	636
QY	1667	CAGTGTAAATACCAACCCCTGTGTATGGAGCAAGTCCCTAACCCGCGAAAGCTAAAC	1728
Db	637	HisSerGluArgIHisValLeuLeuTyGlyThrAspProLeuSerArgTyLeuIAsp	656
QY	1727	AGCAAGAGTGTGTCGGCCATCGACAGCAAAATATGATGTGTCGTGGTGGGGG	1786
Db	657	AsnGlyIHisLeuAlaGlyLeuLeuAlaGlnIAspGlyMetPheValMetLeuProCysIleIly	676
QY	1787	ATCGGGGTGGACATGTGACACCTGTGTGGAAAGAAATACATCCTGTCCAGTCTGTGGCATG	1846
Db	677	ValGlyValaSerThrAspThrIleThrArgGlnCysLeuIHisSerSerAsnAlaIalaArgTrp	696
QY	1847	CGAGTGGGCTGCTGTGGAGACTGGCTTAAAGTGGCTGCGAGAGACTCAAGATGGA	1906
Db	697	AlaAlaGlySerValIThrAspLeuAlaPheLysValAlaIAspArgGluIleuLysAsnGly	716

[illegible]

KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 XX
 XX W0200157190-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001WO-US04098.
 XX
 XX 03-FEB-2000; 2000US-0496914.
 XX 27-APR-2000; 2000US-0560875.
 XX 20-JUN-2000; 2000US-0598075.
 XX 19-JUL-2000; 2000US-0620325.
 XX 01-SEP-2000; 2000US-0654936.
 XX 15-SEP-2000; 2000US-0663561.
 XX 20-OCT-2000; 2000US-0693325.
 XX 30-NOV-2000; 2000US-0728422.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AU, Yang Y, Wejhrman T, Goodrich R;
 XX
 XX WPI: 2001-476283/51.
 DR N-PSDB; AAK53008.
 XX
 XX Claim 20; Page 369-370; 6221pp; English.
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX
 XX
 CC The invention relates to polynucleotides (AAK51456-AAK51435) and the
 CC encoded polypeptides (AAM8323-AAM8302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM80020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 XX
 XX Sequence 1020 AA:
 SQ
 Alignment Scores:
 Pred. No.: 3,98e-115 Length: 1020
 Score: 1831.50 Matches: 424
 Percent Similarity: 56.03% Conservative: 124
 Best Local Similarity: 43.35% Mismatches: 249
 Query Match: 34.26% Indels: 182
 DB: 22 Gaps: 22
 US-09-502-945-2 (1-2885) x AAM79875 (1-1020)
 QY 11 TTGTGCAAGTCAAGAGGCCACACAGGGGCTCAACATTCCTCCACAGACACCCC 70
 DB 169 LenuLnuArgrValrHlSProAnsSerProGlyLeuPyrArgrHlPLeuGluProLeu 188
 QY 71 AAATGCGGGGAGGCCACACATGCTTTTGGACAGAGTTCCCTCCACAGAGGGCCCC 130
 DB 189 GlnhrGlnGlyAlaThrArgrSerMetLeuSerSerPheLeuProProValProSerLeu 208
 QY 131 CTTGGAGAGGCTCCCTTACAAATGCTTTGCTGGGCCCTTACAGACAGTGCAGAGAC 190
 DB 209 ProSerAspProProGluHls----- 215
 QY 191 TTCCCTCGCGCAAAACAGGCTGTGAACCAACTTGAAGAGTGCCTTCAAGGCTAAAAACAG 250

DB 216 PheProLeuArgrLysThrValSerGluProbsnLeuArgrTyrLysProLys--- 234
 QY 251 AAGTGGCTGAGCGAGAGACGATCCCTCGCTGCGACAGATGGAGCTATTATGAC 310
 DB 235 LysSerLeuGlnArgrLysAsnProLeuArgrLysGlnSerAlaPro-----Pro 252
 QY 311 ACCTTTAAGAGAGAGCTGTGAGATCATCAGAGTCCCGGCGCTGGGGCTGCTCGTGTGT 370
 DB 253 SerLeuArgrArgrProAlaGlnThrLeuLysAspSerProSerSer----- 270
 QY 371 AACAGCGACCCGCGCTCGCGCCAGCTCTCCCAACAGCTCCACACACACATGCTGAG 430
 DB 271 ---SerThrProAlaSerGlyCysSerSerProbsnAspSerGluHls----- 285
 QY 431 AATGGCTTTACTGGCTCAGTCCCAACATCCCACTGATGCTCCCTCAGACCGAGCC 490
 DB 286 -----GlyProAsnProIleLeuGlnSerGlnAlaLeuLysGlnArgrLeu 301
 QY 491 CTCCCTCGAGACAGCTCCCGCCACACAGTTCAGCTCTACAGCTCTCTGCTGCGCCAC 550
 DB 302 ArgLeuGlnGlnThrSerValAlaProPheAlaLeuProThrValSerLeuLeuProAla 321
 QY 551 ATCTCCCTAGGGCTGACAGGCCACGCTCATGTCACCACTCACACTGCTGCTCCCG 610
 DB 322 IlleThrLeuGlnLeuProAla----- 328
 QY 611 AAGCTGTGCACACAGAGAGGCCGAGAGAGCGGCTCCAGTCCGCGGACAGGTGCG 670
 DB 329 -----ProAlaArgrAlaAspSerAspArgrHrHlsProThrLeuGlnProArgrGly 346
 QY 671 ACGGTACCGGGCAGGTTCATGACACATCCCTTCCTGCTGCTGCGGTGGCGGCA 730
 DB 347 ProIleLeuGlnSerProHlsThrProLeuPheLeuProHls-----GlyLeuGlu 363
 QY 731 CTGGAGGCGCAGCGAGAGCCCGCCAGGCAATGCTCCCTGCTGACATGATGCTTGTCTG 790
 DB 364 ProGluAlaGlyGlyThr-----LeuProSerArgrLeuGlnProIleLeuLeu 380
 QY 791 GAGAGGCGCGGCGACAGAGACCCCTATTGCTGTGCTCCACTCCAGCGGCAAGTCCCACTA 850
 DB 381 AspProSerGlySerHlsAlaProLeuLeuThrValProGlyLeuGlyProLeuProPhe 400
 QY 851 GTGAGGGGTGAACGTGTGGCCACCGACATGCGAGCGGTAGACAGTCCCGGCGCATCGG 910
 DB 401 HisPheAlaGlnSerLeuMetThrThrGlnArgrLeuSerGlySer---GlyLeuHlsTrp 419
 QY 911 CCCCTGAGCGGCACTCAGTCCCTCAACGCTGCGCGCAGAGTCCCGAGGCC----- 958
 DB 420 ProLeuSerArgrHrArgrSerGlnProLeuProProSerAlaThrAlaProProPro 439
 QY 959 -----CTCCAGACACTGTGTATGCAACAACAGCAGCCAGCTTCTCTG 1000
 DB 440 GlyProMetGlnProArgrLeuGlnGlnLeu-----LysThrHlsValGlnValIle 456
 QY 1001 GAGAGCAGAGAAGCAGCAGACGCTACAGCTGGGCAAGATCTCAACACAGAGAGGAGACTG 1060
 DB 457 LysArgrSer-----AlaLysProSerGluLys 465
 QY 1061 CCCAGCAGGCCACACACACACCTGAGAGAGACAGAGAGAGTGAAGGAGCAGAGAGAG 1120
 DB 466 ProArgr----- 467
 QY 1121 GTCTTGCTGGGGAGGAGGCCCTGACCATGCCCCGGAGAGGCTCCACAGAGTGAAGC 1180
 DB 467 ----- 467
 QY 1181 ACACAGAGAAGACTGAGAGAGAGAGAGAGAGATGGAGAGAGAGATGTC 1240
 DB 468 LeuArgrGlnIleProSerAlaGlnAspLeuGlnThrAspGly----- 481
 QY 1241 ATCCAGGTTAAGACGAGAGAGGCGAGAGTGTG-----GCTGAGAGAGGGCGCGACTTG 1294

Db 482 -----glylyProglyValValAspaspolyLeuGlnHis 494
 QY 1295 GAGGAGCCTGTGCTGTGATACAAAAAAGTGTCTCAGATGCCCAACGGCTTGCAACCTTTG 1354
 Db 495 ArgGlyLeuGlyHisGlyIn -----ProGlnAlaArgGlyProAlaProLeu 510
 QY 1355 CAGGTGTACCAAGCGCC----- 1372
 Db 511 Gln-----GlnHisProGlnValLeuLeuTrpGlnGlnArgLeuAlaGlyArgLeu 528
 QY 1373 -----CTCAGCCTGGCCCACTGTGCCCAACCAAGCC 1402
 Db 529 ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaGlnGlyValHisArgPro 548
 QY 1403 CTGGGCGCTACCAACATCTCCCTGCTGCCCTGGGGGATGAAAGAACCC----- 1453
 Db 549 LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerAlaProGlnProAla 568
 QY 1454 -----CCAGAACACCCGTCAGACCTC---TTCACCCACA 1486
 Db 569 SerGlnAlaArgValLeuSerSerSerGlnThrProAlaArgThrLeuProPheThrThr 588
 QY 1487 AGTGTGCTGTACGACACCTTCATGCTAAAGCAACCACTGCATGTGGCGGAACACACAGCTG 1546
 Db 589 GlyLeuIleTyrAspSerValIleMetLeuLysHisGlnCysSerGlyAspAsnSerArg 608
 QY 1547 CACCCGTGACATGCTGGCGGATCCAGACATGCTGCCGCTGCAGAGACAGCCTG 1606
 Db 609 HisProGlnHisAlaGlyArgIleGlnSerIleTrpSerArgLeuGlnGlnAlaGlyLeu 628
 QY 1607 CTTAGCAATGCGAGCGGATCCGAGGTGCGAAAGCCACGCTAGAGATCCAGACAGT 1666
 Db 629 ArgSerGlnCysGlnCysLeuArgGlyArgLysAlaSerLeuGlnGlnLeuGlnSerVal 648
 QY 1667 CACTGTGATACACACACCTGCTGTATGGGACAGCAGCTCCCAACCGGAGAAAGCTAGAC 1726
 Db 649 HisSerGlnArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysLeuAsp 668
 QY 1727 ACCAAGAAAGTGTGCTGCCATCCATCAGACCGAAGATGATGTGTGCTGTGGGGC 1786
 Db 669 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGlyGly 688
 QY 1787 ATCGGGGTGACAGTACACACCGCTGTGAATGAGATGCACTCTCCAGTCTGTGCGCATG 1846
 Db 689 ValGlyValAspThrAspThrIleTrpAsnGlnLeuHisSerSerAsnAlaAlaArgTyr 708
 QY 1847 GCAGTGGCTGCTGCTGTGAGCTGGCTTCAGAGTGGCTGCAGAGAGCTCAAGATGGA 1906
 Db 709 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGlnLeuLysAsnGly 728
 QY 1907 TTGGCATATCCGGCCCGCAGACACACCGCGAGGATCCACAGCCTATGGATGTGCG 1966
 Db 729 PheAlaValValArgProProGlnHisHisAlaAspHisSerThrAlaMetGlyPheCys 748
 QY 1967 TTCTCACTGTGTAGCCATCAACCGCAAAATCTCTACAGCAGAGATTTGAACGTGGGCAAG 2026
 Db 749 PhePheAsnSerValAlaIleAlaLysArgGlnLeuGlnGlnIleSerLysAlaSerLys 768
 QY 2027 GTCCATCATGTGAGTGGGACATTCACCATGGCAATGGGACACCGAGCGTTCTCATAT 2086
 Db 769 IleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnGlnInThrPheTyrGln 788
 QY 2087 GACCCCTGTGCTGTATCTCTGTGATCGTATGATCAACAGGGAATCTTTTCCAGGC 2146
 Db 789 AspProSerValLeuTyrIleSerLeuHisArgHisAspAspLysAsnPhePheProGly 808
 QY 2147 TCTGGGGCTCTGAAGAGTGTGGAGAGCAGCGGTGGGTATCAATGTGAACGTGGA 2206
 Db 809 SerGlyAlaValAspGlnValGlyAlaGlySerGlyGlnGlyPheAsnValAsnValAla 828
 QY 2207 TGGACAGGAGGTGTGAGACCCCGCATGTGAGACGTGAGATACCTTACACCTTCAGAGACA 2266
 Db 829 TrpAlaGlyGlyLeuAspProProMetGlyAspProGlnTyrLeuAlaAlaPheArgIle 848

QY 2267 GTGTGATGCCATTTGCCAGAGATTCTCAGTATGTGTCTTACCTCCGGCGGTTT 2326
 Db 849 ValValMetProIleAlaArgGlnPheSerProAspLeuValLeuValSerAlaGlyPhe 868
 QY 2327 GATGCTGTGAAGAGATGTGTCTCTGTGGGTGCTTACTGTCTGACCCAGATGTTTT 2386
 Db 869 AspAlaIleGlnGlyHisProAlaProLeuGlnGlyTyrHisValSerAlaLysCysPhe 888
 QY 2387 GGGCACTTGACAGCAGCTGATGACCTGGGAGGGGCGGGGTGTGGTGGCGCTGGAG 2446
 Db 889 GlyTyrMetThrGlnGlnLeuMetAsnLeuAlaGlyGlyAlaValValLeuAlaLeuGln 908
 QY 2447 GGAAGCCATGACTGACCCCGCATCTGTGATGCTGTGAAGCTTGTCTGCGCTGTCTC 2506
 Db 909 GlyGlyHisAspLeuThrAlaIleCysAspAlaSerGlnAlaCysValAlaAlaLeuLeu 928
 QY 2507 AGTGTAACTGACGCGCTTGGATGAGCAGTCTTCACGAAAGCCCAACATCAACGGA 2566
 Db 929 GlyAsnArgValAspProLeuSerGlnGlnGlyTyrTrpLysGlnLysProAsnLeuAsnAla 948
 QY 2567 GTGGCCAGCCTGAGAAAGTCATGAGATCCAGACGCAACACTGAGCTGTGTGCGAAG 2626
 Db 949 IleArgSerLeuGlnAlaValIleArgValHisSerLysTyrTrpLysMetGlnArg 968
 QY 2627 TTGCGCGTGTCTGTGGCGCGTCCCTGCGAGGGGCCCAAGCAAGTGAACCGAAGCC 2686
 Db 969 LeuAlaSerCysProAspSerTrpValProArgValProGlyAlaAspLysGlnGlnVal 988
 QY 2687 GAAT- GTGACAGCCATGGCTTGTGTGTGGG-----GCCAACAAGGCCCAAGCT 2739
 Db 989 GlnAlaValThrAlaLeuAlaSerLeuSerValGlyIleLeuAlaGlnLysP----- 1005
 QY 2740 GCGGAGCCCGGGAACACAGCCCGGCGGAGAGAGCCCATGTGAGCAGAGAG 2793
 Db 1006 -----ArgProSerGlnGlnLeuValGlnGln 1015
 RESULT 10
 ID AA97033 standard; Protein; 1428 AA.
 XX AA97033;
 DT 31-OCT-2000 (first entry)
 XX
 DE Caspase 8-interacting protein from PAC RPC15-1057120.
 XX
 KW Caspase-8 interacting protein; CIP; caspase-8 inhibitor; anti-apoptotic;
 KW N-acetylglycosamine-6-phosphate deacetylase; p35 TNF-R; neuroprotective;
 KW tumour necrosis factor receptor; anti-inflammatory; anti-diabetic;
 KW hepatotropic; virucide; cardiant; anti-ulcer; chromosome 12q31.
 XX
 OS Homo sapiens.
 XX
 PN WO200039160-A2.
 XX
 PD 06-JUL-2000.
 XX
 PE 23-DEC-1999; 99WO-IL00698.
 XX
 PR 24-DEC-1998; 98IL-0127721.
 PR 28-SEP-1999; 99IL-0132105.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 PI Wallach D, Schuchmann M, Goncharov T;
 DR WPI, 2000-452371/39.
 XX
 PT New caspase-8 interacting protein capable of interacting with subunit 1
 PT and/or subunit 2 of caspase-8, useful in the treatment of multiple
 PT sclerosis with primary oligodendroglipathy, autoimmune uveoretinitis
 PT and diabetes

XX Claim 3; Fig 6; 125bp; English.

CC A caspase-8 interacting protein (CIP), capable of interacting with
 CC subunit 1 and/or 2 of caspase-8, was identified by a yeast two hybrid
 CC screen in human placental cDNA clone P74. A similar sequence was also
 CC found in a genomic clone (RPC15-1057120) which localizes to human
 CC chromosome 12q31. Clone P74 shows homology to a highly homologous region
 CC of mouse and human histone deacetylases. CIPs inhibit caspase-8 and human
 CC p53 tumour necrosis factor receptor (TNF-R) induced apoptosis in HEK
 CC 293-T and HeLa cells. The CIP, CIP-targeted ribozymes, antisense
 CC oligonucleotides or antibodies are useful for modulating caspase-8
 CC activity. TNF-R or Fas-mediated effects, or apoptosis. They are also
 CC useful in the treatment of multiple sclerosis with primary
 CC oligodendroglopathy, autoimmune uveoretinitis, diabetes, lupus,
 CC autoimmune myocarditis I, hepatitis C virus (HCV) mediated chronic
 CC hepatitis, chronic gastritis (e.g. type A gastritis), mixed connective
 CC tissue disease (MCTD), Crohn's disease, or ulcerative colitis. CIP are
 CC also useful for the isolation, identification and cloning of other
 CC proteins of the same class.

XX Sequence 1428 AA:

Alignment Scores:
 Pred. No.: 7.02e-109 Length: 1428
 Score: 1739.50 Matches: 423
 Percent Similarity: 50.75% Conservative: 121
 Best Local Similarity: 39.46% Mismatches: 239
 Query Match: 32.54% Indels: 289
 DB: 21 Gaps: 26

US-09-502-945-2 (1-2885) x AAY97033 (1-1428)

QY 21 CAAGGAGGCCACACAGGGGCGCTCAACCTTCCTCCACAGCACC---CCAATGCT 77
 Db 400 GUAtrgThrValAlaProAsnSerProGlyLeuProTyArgSerGlnGlyProCysSer 419
 QY 78 GGGAGGCCACCATCTTCTTGGACCAAGATTCCCTCCACAGAGGGCCCTCGGGA 137
 Db 420 GLyIn-CysProCys-----SerValProThrProLeuLysGlnProTPrHl 435
 QY 138 CGCC-----TC 143
 Db 435 sSerPheCysArgThrLeuGluProLeuGluThrGluGlyAlaThrArgSerMetLeu 455
 QY 144 CCGCTCAACAACTGCGCTTGGCTGG---CCCTACAGACAGCGAGAGACTTCCCTCC 200
 Db 455 r-SerPheLeuProProValProSerLeuProSerAspProProGlnHisPheProLeuA 475
 QY 201 GCAAAAGAGCCTCGAACCCCAACTTGAAGTGCCTCAAGGCTAAACAGAGGTGGCTG 260
 Db 475 rGlySerThrValSerGluProAsnLeuLysLeuAlaGlyTyrLysProLys---LysSerLeuG 494
 QY 261 AGCGAGAGACAGTCCCTCCCTGCGTGCAGAGATGGAGACTGTTATTAGCACTTTAGA 320
 Db 494 lnaArgArgLysAsnProLeuLeuArgLysLysSerAlaPro-----ProSerLeuArgA 512
 QY 321 AGAGAGCTGTGAATACACAGTCCGGGCGCTGGGGCGTGCCTGCTGTAAACGGCCAC 380
 Db 512 rGAGGProAlaGluThrLeuLysPserSerProSerSerSer-----SerThrP 529
 QY 381 CGGCGTCCGGCCCGCTCCCAACAGCTCCACAGCTCCACAGACATGCTGAGAAATGCTTTA 440
 Db 529 roAlaSerGlyCysSerSerProAsnAspSerGlnHis-----541
 QY 441 CTGGCTCAGTCCCAACATCCCACTGAGATGCTCCCTCAGACAGCGCCCTCCCTG 500
 Db 542 --GlyProAsnProIleLeuGlySerGlnAlaLeuLeuGlyGlnArgLeuArgLeuGlnG 561
 QY 501 ACAGCTCCCCACCACTAGCCTTACACAGTCTCTTCTGTGACCAATCTCCCTAG 560
 Db 561 luthrSerValAlaProPheAlaLeuProThrValSerLeuLeuProAlaIleThrLeuG 581

QY 561 GGCTGACAGCCAGCGTCACTGTGTACACAATCACAACCTCACTGCTCCCGAGCTGTGCA 620
 Db 581 lyleuProAla-----ProA 586
 QY 621 CACAGCAGAGGCCGAGAGCAGGCGCTCCAGTCCCTGCGGAGGCTGACGCTGACCG 680
 Db 586 lalArgAlaAspSerAspArgArgThrHisProThrLeuGlnProArgGlyProIleLeuG 606
 QY 681 GCAAGTTCATGAGACACTCTCTATTCCTGGCTGCGCTGCGGCGGTGGACCTGAGGGCG 740
 Db 606 lyleuProIsthTrProLeuPheLeuProHis-----GlyLeuGlnProGlnAlaG 623
 QY 741 ACAGGAGCCCGCCACAGGAGATGCTCCCTGCTGACAGATGCTGTGCTGAGACAGCGCC 800
 Db 623 lyleuThr-----LeuProSerArgLeuGlnProIleLeuLeuLeuAspProSerG 640
 QY 801 GGCAGCAGACACCTCATTTGCTGTGTGCTCCACAGGCGAGTCCCACTAGTACGGGTG 860
 Db 640 lyleuHisAlaProLeuLeuThrValProGlyLeuGlyProLeuProPheHisPheAlaG 660
 QY 861 AACGTGGGCGCCACAGATGCGGAGCGTAGGCAAGCTCCCGGCGATCGGCGCTGAGCC 920
 Db 660 lyleuMetThrThrGluArgLeuSerGlySer---GlyLeuHisThrProLeuSerA 679
 QY 921 GCACTCAGTCCCTGACCGCTGCGGAGAGTCCCGGAGCC-----958
 Db 679 rGlyThrArgSerGluProLeuProProSerAlaThrAlaProProProProGlyProMetG 699
 QY 959 -----CTGCACAGCTGTGTCATGCACACAGCAGCAGAGCTTCTGAGAGAGCAGA 1010
 Db 699 lnpProArgLeuGlnGlnLeu-----LysThrHisValGlnValIleLysArgSer- 715
 QY 1011 AGCAGCAGCAGCTACACTGGGCAAGATCCTACCAACAGACAGGGGAGCTCCGACAGC 1070
 Db 716 -----AlaLysProSerGlnLysProArg---723
 QY 1071 CCACACCCACACCTGAGAGACAGAGAGAGAGAGCTGACGAGCAGCAGAGCTTGTGCG 1130
 Db 723 -----723
 QY 1131 GGGAGGAGCCCTGACCATGCGCCCGGAGGGCTCCACAGAGAGTGAAGCAGCAGAGAG 1190
 Db 724 -----LeuArgGlnI 727
 QY 1191 ACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTCAGTCCAGGTGA 1250
 Db 727 lerpSerAlaGluAspLeuGluThrAspGly-----737
 QY 1251 AGGACGAGAGAGGCGGAGAGTGT-----GCTGAGAGAGGGCGCCGACTTGGAGAGCTG 1304
 Db 738 -----GlyGlyProGlyGlnValAlaAspArgLysLeuGlnHisArgGlyLeuG 754
 QY 1305 GTGCTGATACAAAAAAGTGTCTCAGATGCCCAACCGGTGCAACCTTTCAGAGTGTAC 1364
 Db 754 lnhHisGlyGln-----ProGlnAlaArgGlyProAlaProLeuGln-----G 768
 QY 1365 AAGCGCCC-----1372
 Db 768 lnhIsProGlnValLeuLeuTrpGlnGlnAlaGlyAlaGlyArgLeuProArgLys 788
 QY 1373 -----CTCAGCTGGCCACTGTGCGCCCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1412
 Db 788 erThrGlyAspThrValLeuLeuProLeuAlaGlnGlyGlnHisArgProLeuSerArgA 808
 QY 1413 CCACATCTCCCTGCTGCGCTGGGGGCGATGAAGAGCC-----1453
 Db 808 lalInSerSerProAlaAlaProAlaSerLeuSerAlaProGluProAlaSerGlnAlaA 828
 QY 1454 -----CCAGACCAACCCCTCAAGACACTC---TTCACCAACAAGTGTGTC 1496
 Db 828 rValLeuSerSerSerGlyThrProAlaArgThrLeuProPheThrGlyLeuIleT 848
 QY 1497 ACGACAGCTTCATGTAAACACACAGTGTGTGGGAGACACACAGTGCACCTGAGC 1556

```

|||||: ||||||| ||||||| |||||||: |||||||
Db yAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArgHisProGlnH 868
848 yAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArgHisProGlnH 868
OY ATGCGGGCGGATCCAGACCATCTGGTCCGCGAGAGACAGCCGCTTAGCAACT 1616
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db lAlaGlyArgTleGlnInsrtLleTrrPserArgLeuGlnLysArgLysLeuArgSerGlnC 888
868 lAlaGlyArgTleGlnInsrtLleTrrPserArgLeuGlnLysArgLysLeuArgSerGlnC 888
OY GCGACCGATCCGAGTCCGAAGCCACGCTAGATGATGCCAGACAGTGCATGTGAAT 1676
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db yGlnCysLeuArgGlyArgGlyAlaSerLeuGlnLysLeuGlnInsrtValHisSerGlnA 908
888 yGlnCysLeuArgGlyArgGlyAlaSerLeuGlnLysLeuGlnInsrtValHisSerGlnA 908
OY ACCAGACCTGCTATGAGGACAGTCCCTCAACCGGCAAGACTAGACAGCAAGACT 1736
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db rGHisValLeuLeuArgTyrGlnHisnProLeuSerArgLeuLysLeuAspAsnGlyLysL 928
908 rGHisValLeuLeuArgTyrGlnHisnProLeuSerArgLeuLysLeuAspAsnGlyLysL 928
OY TGCCTGGTCCCATCCAGCCAGAGATGATGCTGCTGCTGGGCGCATCCGGG----- 1792
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db eAlaGlyLeuLeuAlaGlnArgMetCpHeValMetLeuProCysGlyGlyValGlyProL 948
928 eAlaGlyLeuLeuAlaGlnArgMetCpHeValMetLeuProCysGlyGlyValGlyProL 948
OY ----- 1792
1792 ----- 1792
Db eAlaThrLeuSerAlaPheLeuAlaSerLeuAlaProThrValProGlnGlyLeuSerA 968
948 eAlaThrLeuSerAlaPheLeuAlaSerLeuAlaProThrValProGlnGlyLeuSerA 968
OY ----- 1792
1792 ----- 1792
Db rGValSerTrpGlyLeuLysProProProGlyProAsnProLysSerArgProAlaProC 988
968 rGValSerTrpGlyLeuLysProProProGlyProAsnProLysSerArgProAlaProC 988
OY ----- 1792
1792 ----- 1792
Db ySPTrpGlyProGlyArgGlyValGlyThrThProLeuGlyProGlySerCysValL 1008
988 ySPTrpGlyProGlyArgGlyValGlyThrThProLeuGlyProGlySerCysValL 1008
OY ----- 1792
1793 ----- 1793
Db ySPTrpMetMetArgAlaLeuThrLeuAlaProGlnValAspTrpHisArgTrpLleTrrP 1028
1008 ySPTrpMetMetArgAlaLeuThrLeuAlaProGlnValAspTrpHisArgTrpLleTrrP 1028
OY ATGACATGACATCCCTCCAGTGCCTGCGCGATGCGAGTGGCGCTGCGAGCTGGCT 1874
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db sGlnLeuHisSerSerAsnAlaAlaArgTrrPAlaAlaGlySerValLTrpAspLeuAlaP 1048
1028 sGlnLeuHisSerSerAsnAlaAlaArgTrrPAlaAlaGlySerValLTrpAspLeuAlaP 1048
OY TCAAGGTGGCTGCAGAGAGCTCAAGAAATGATGATTTGCCATCCGCGCGCGAGAGACCC 1934
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db hElyValAlaSerArgGlyLeuLysAsnGlyPheAlaValAlaArgProProGlyHisn 1068
1048 hElyValAlaSerArgGlyLeuLysAsnGlyPheAlaValAlaArgProProGlyHisn 1068
OY ACGCGAGAGAAATCCACAGCCATGGATTCGCTTTCACCTGTAGCCATCACCGCAA 1994
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db lAlaAspHisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysA 1088
1068 lAlaAspHisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysA 1088
OY AACTGCTACAGACAAAGTGAACGTGGGCAAGTCTCTATGCTGGACTGGACATTCAAC 2054
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db rGlnLeuGlnGlnInsrtLysAlaSerLysIleLeuIleValAspTrpAspValHisn 1108
1088 rGlnLeuGlnGlnInsrtLysAlaSerLysIleLeuIleValAspTrpAspValHisn 1108
OY ATGGCAATGGCAGCCAGCGGCTTCTACAAATGACCCCTGCGCTGCTACATCTCTGCG 2114
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db lSGlYAsnGlyThrGlnGlnInsrtPheTyrGlnAspProSerValLeuLysIleSerLeuH 1128
1108 lSGlYAsnGlyThrGlnGlnInsrtPheTyrGlnAspProSerValLeuLysIleSerLeuH 1128
OY ATCGCTATGACAAACGGAATCTTTCACAGCTCTGGGCTCTGAGAGAGTGTGTGAG 2174
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db lAspArgHisAspArgLysAsnPheProGlySerCylValAlaValAspGluValGlyAla 1148
1128 lAspArgHisAspArgLysAsnPheProGlySerCylValAlaValAspGluValGlyAla 1148
OY GACCAAGCGCTGGGCTACATGTGAACGTGGCATGGACAGAGAGTGGAGCCCCCATATG 2234
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db lYSerGlyGlnGlyPheAsnValAsnValAlaIleTrrPAlaGlyLeuAspProPromeG 1168
1148 lYSerGlyGlnGlyPheAsnValAsnValAlaIleTrrPAlaGlyLeuAspProPromeG 1168
OY GAGAGGTGAGTACCTTACAGCTTCAGAGACAGTGTGATGGCCATGGCCAGAGTTCT 2294
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db lYAspProGlyLysLeuAlaAlaPheArgIleValAlaMetProIleAlaArgGluPheS 1188
1168 lYAspProGlyLysLeuAlaAlaPheArgIleValAlaMetProIleAlaArgGluPheS 1188
OY CACCTGATGTGCTCTAGTCTCCGCGGCTTGAATGCTGTGTAAGACATCTGTCTCTC 2354
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db ePProAspLeuValLeuValSerAlaGlyPheAspAlaAlaGlnGlyHisProAlaProL 1208
1188 ePProAspLeuValLeuValSerAlaGlyPheAspAlaAlaGlnGlyHisProAlaProL 1208
OY TGGGTGGCTACTCTGTACACCGCAGATGTTTGGCCACTTGACACAGGACGCTGATGACCC 2414
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```

```

Db 1208 eUGlyGlyTyrHisValSerAlaLysCysPheGlyTyrMetThrGlnGlnLeuMetAsnL 1228
OY 2415 TGCAGAGGCGCGGCTGCTGCTGCGAGAGAGCCATGACTTGAACGCCATCTGTG 2474
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1228 eAlaGlyLysAlaValIleAlaLeuGlnLysGlyHisAspLeuThrAlaIleCysA 1248
OY 2475 ATGCTCTGAACCTTGTCTGCTGGCTCTGCTCAAGTGAAGCTGACGCCCTTGATGAG 2534
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1248 sPAlaSerGlnAlaCysValAlaAlaLeuLeuGlnLysAsnArgValAspProLeuSerGlnG 1268
OY 2535 CAGCTTCGACGAAAGCCCAACATCAACAGCACTGGCCACCGCTAGAGAAATCATCGA 2594
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1268 lUGlyTrrPysGlnLysProAsnLeuAsnAlaIleArgSerLeuGlnAlaValIleArgV 1288
OY 2595 TCCAGAGCAACACATGACCTGTGTGCAAGATTCGCCCTGGTGTGGCGGCTCCCTGC 2654
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1288 aHisSerLys----- 1291
OY 2655 GAGGGGCCCAACAGCTGAGACCGAAGCCGAATGTGAAGCCATGCGCTTGTCTGT 2714
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1292 -----CysGlyAspGlyThrLeuAlaGlnLeu----- 1300
OY 2715 GGTGGGGCCGAACAGCCCAAGCTGCGCAGCCCGGGAACACAGCCCGGAGAG 2774
|||||: ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1301 -----ArgLeuLysAspLeuGlyLysThrLeuProHisArgGlyG 1314
OY 2775 GGAGCCCATGAGACAGACCTGCCCCCTGTGAC 2806
Db 1314 lInleuGlyPheArgCysGlnProGlyAsp 1324

RESULT 11
AAU99662
ID AAU99662 standard; Protein; 855 AA.
AC AAU99662;
XX
XX 07-OCT-2002 (first entry)
DE Human Histone deacetylase Isoform 7.
XX
XX Human; histone deacetylase; HDAC-7; cancer; cytostatic;
KW antisenese; tumour suppressor; cell proliferation; tumour;
KW programmed cell death; necrotic cell death.
XX
XX Homo sapiens.
OS
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 54 /note= "Encoded by ACG"
FT FT Misc-difference 117 /note= "Encoded by CCG"
FT FT Misc-difference 120 /note= "Encoded by CCG"
FT FT Misc-difference 134..135 /note= "Encoded by CTTCCG"
FT FT Misc-difference 137 /note= "Encoded by GCT"
FT FT Misc-difference 140 /note= "Encoded by CCG"
FT FT Misc-difference 148 /note= "Encoded by TTC"
FT FT Misc-difference 163 /note= "Encoded by CAT"
FT FT Misc-difference 169 /note= "Encoded by CCG"
FT FT Misc-difference 186 /note= "Encoded by CCC"
FT FT Misc-difference 211 /note= "Encoded by CCC"
FT FT Misc-difference 261 /note= "Encoded by ACC"
FT FT Misc-difference 270 /note= "Encoded by CCT"

```


Db 324 SerAlaThrAlaProProProGlyProMetGlnProArgLeuGlnLeu----- 341
OY 977 CAACAACAGCAGCAGTCTCTGGAGAAGACAGACAGCAGCAGTACAGCTGGGCAAG 1036
Db 342 ---LysThrHisValGlnValIleLysArgSer----- 351
OY 1037 ATCTCCACCAAGCAGGAGCTGCCAGGACCCACCAACCCTGAGGACAGAG 1096
Db 352 -----AlaLysProSerGlnLysProArg----- 359
OY 1097 GAGGAGCTGACGAGCAGCAGGAGCTGTCTGGGGAGGAGCCCTGACATGCCCGG 1156
Db 359 ----- 359
OY 1157 GAGGGCTCCACAGAGTAGAGACACAGAGACCTGGAGGAGGAGCAGCAAGAG 1216
Db 360 -----LeuArgGlnIleProSerAlaGlnLysProLeuGln 371
OY 1217 GATGGGAGGAGGAGGAGGAGTTCATTCACAGTTAAAGCAGGAGGCGCAGAGTGGT--- 1273
Db 372 AspGly-----GlyLysProGln 378
OY 1274 ---GCTGAGAGGAGGCGCCGACTTGGAGAGCCTGGTGTGATACAAAAGTGTCTCA 1330
Db 379 ValValAspAspGlyLeuGlnHisArgGlnLeuGlnHisGlyGln-----Pro 394
OY 1331 GATGCCCAACCGCTGCACCTTTCAGGTGTACCAAGCGCC----- 1372
Db 395 GlnAlaArgGlyProAlaProLeuGln-----GlnHisProGlnValIleIleTrpGln 412
OY 1373 -----CysCAGC 1378
Db 413 GlnGlnAlaArgLeuAlaGlyArgLeuProArgGlySerThrGlyAspCysValIleLeuPro 432
OY 1379 CTGGCACTGTGTCCCAAGCCCTGGGCGCCTACCCATCTCCCTGGCTGGG 1438
Db 433 LeuAlaGlnGlyGlnHisArgProLeuSerAlaGlnSerProAlaAlaProAla 452
OY 1439 GGCATGAGAACCC-----CCAGACCAACCC 1465
Db 453 SerIleSerAlaProGlnProAlaSerGlnAlaArgValLeuSerSerSerGlnTrpPro 472
OY 1466 GTCAAGCACCTC---TTCACCAAGTGTGTCTACGACACCTGATCTGTAACCAACCG 1522
Db 473 AlaArgThrLeuProPheLeuThrGlyLeuIleTyrAspSerValIleLeuLysHisGln 492
OY 1523 TTCATGTGGGAGACACACACCTGACACCTGACATGTGTGCGCGATCCAGACATCTGG 1582
Db 493 CysSerCysGlyAspAsnSerArgHisProGlnHisAlaGlyArgIleGlnSerIleTrp 512
OY 1583 TTCGCGCTGCAGGAGACAGCCTGCTTACAAAGTGCAGGCGGATCCGAGGTGCACAGCC 1642
Db 513 SerArgLeuGlnGlnArgGlyLeuArgSerGlnCysGlnCysLeuArgGlyArgLysAla 532
OY 1643 AGCTAGATGAGATCCAGACATGTGCATCTGAATACCAACCTGTCTATGGAGCCAGT 1702
Db 533 SerIleGlnGlnLeuGlnSerValHisSerGlnArgHisValLeuLeuTrpGlyThrAsn 552
OY 1703 CCCCTCAACCGGACAGATAGACAGAAAGTTGCTGCTCCATCCATGAGCCAGAGATG 1762
Db 553 ProLeuSerArgLeuLysLeuAspAsnGlnLysLeuAlaGlyIleIleLeuAlaGlnArgMet 572
OY 1763 TATGCTGTGCTGCTTGTGGGGGATCGGGGTGACAGTGCAGACCGTGTGAATAGATG 1822
Db 573 PheGlnMetLeuProCysGlyGlyValGlyValAspThrAspThrIleTrpAsnGlnLeu 592
OY 1823 CACTGCTCCAGTGTGTGGCATGTGCAGTGGGCTGCTGCTGAGCTGAGCTTCAAGGTG 1882
Db 593 HisSerSerAsnAlaAlaArgTrpAlaAlaGlySerValThrAspLeuAlaPheLysVal 612
OY 1883 GCTGAGAGGAGCTCAAGATGATTTGGCATCCGCGCCCGAGCAACACAGCCGAG 1942
Db 613 AlaSerArgGlnLeuLysAsnGlyPheAlaValValArgProGlnHisHisAlaAsp 632

OY 1943 GAATCCACAGCCATGGGATTTGCTTCTTCAACTGTGTAGCCATCCGCAAACTCTTA 2002
Db 633 HisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIleAlaCysArgGlnLeu 652
OY 2003 CAGCAGAG-----TTGAACGTGGGCAAGTCCCTCATCTGAGTGGAGCATTCAC 2053
Db 653 GlnGlnGlnSerLysAlaSerLysAlaSerLysIleLeuIleValAspTrpAspValHis 672
OY 2054 CATGGCAATGGACCCAGCAGGCGCTTTCATATGACCCCTGTGCTGTACATCTCTGG 2113
Db 673 HisGlyAsnGlyThrGlnIleThrPheTyrGlnAspProSerValLeuTrpIleSerLeu 692
OY 2114 CATGCTATGACACCGGAATCTTTCAGAGCTGTGGGCTCTGAAAGAGTGTGGTA 2173
Db 693 HisArgHisAspAspGlyLysAsnPheProGlnLysSerGlyAlaValAspIleValGlyAla 712
OY 2174 GCACACGCGCTGGGTACATGTGAACGTGCATGACAGAGAGTGTGACCCCGCATT 2233
Db 713 GlySerGlyGlnGlyPheAsnValAlaIleAlaTrpAlaGlyGlyLeuAspProMet 732
OY 2234 GAGAGAGTGTACCTTTCACGCTTTCAGACAGTGTGATGCTCCATTTGCCACGATTC 2293
Db 733 GlyAspProGlnTyrLeuAlaAlaPheArgIleValValMetProIleAlaArgGlnPhe 752
OY 2294 TCACCTGATGTGCTCTAGTCTCCGCGGTTTGAATGCTGTGAAGCATCTGTCTCT 2353
Db 753 SerProAspLeuValLeuValSerAlaGlyPheAspAlaAlaGlnGlyHisProAlaPro 772
OY 2354 CTGGGTGCTACTGTGTACACCCGACAGATGTTTGGCCCACTTTCACAGCAGCTGTAGAC 2413
Db 773 LeuGlyGlyTyrHisValSerAlaLysCysPheGlyTyrMetThrGlnGlnLeuMetAsn 792
OY 2414 CTGCAGGAGGCGCGGTGTGCTGGCCTGAGAGGAGGACATGACTTGAACGCACTGT 2473
Db 793 LeuAlaGlyGlyAlaValValLeuAlaLeuGlnGlyGlnHisAspLeuThrAlaIleCys 812
OY 2474 GATGCTGTGAAGCTTGTGTCTGGCTGTCTGATGTAACCTGACACGCTTGGATGAG 2533
Db 813 AspAlaSerGlnAlaCysValAlaIleAlaLeuLeuGlnGlyAsnArgValAspProLeuSerGln 832
OY 2534 GCAGCTTGCACGCAAAAGCC 2554
Db 833 GlnGlyTyrPheGlnLysPro 839
RESULT 12
AAB92700
ID AAB92700 standard; Protein: 614 AA.
AC AAB92700;
XX 26-JUN-2001 (first entry)
DT
XX
XX
DE Human protein sequence SEQ ID No:11098.
KM Human; primer: detection; diagnosis; antisense therapy; gene therapy.
OS Homo sapiens.
XX
XX
PN EP1074617-A2.
PD 07-FEB-2001.
XX
XX
PE 28-JUL-2000; 2000EP-0116126.
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX
PA (HELI-) HELIX RES INST.
XX

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 XX
 PS Claim 8; SEQ ID 11098; 2537bp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 XX Sequence 614 AA;

Alignment Scores:

Pred. No.: 4,39e-100 Length: 614
 Score: 1607.50 Matches: 343
 Percent Similarity: 61.32% Conservative: 85
 Best Local Similarity: 49.14% Mismatches: 131
 Query Match: 30.07% Indels: 140
 DB: 22 Gaps: 14

US-09-502-945-2 (1-2885) x AAB92700 (1-614)

QY 851 GTGACGGGTGACGTGTGGCCACACGATGCGGACGTAGGCATCCCGCGCATCG 910
 ::::||||| :|||||:|||||: ||| |||
 Db 1 MethThrThrluArgLeuSerGlySer-----GlyLeuHisTrp 13
 QY 911 CCCCTGACGCCGACCTCTCTACCCGCTGCCGAGAGTCCCCAGGCC----- 958
 :|||||:|||||:|||||: ||| |||
 Db 14 ProLeuSerArgThrluArgSerGluProLeuProProSerAlaThrAlaProProPro 33
 QY 959 -----CTGACGAGCGTGGTCAATGCAACACGACACGACGAGTTCCTG 1000
 :|||||:|||||:|||||: ||| |||
 Db 34 GlyProMetGlnProArgLeuGluGlnLeu-----LysThnHisValIleValIle 50
 QY 1001 GAGAGACGAGAGACGACGACGATACAGCTGGGCAAGATCTTCACCCAGAGGAGCTG 1060
 ::::||||| :|||||:|||||: ||| |||
 Db 51 LysArgSer-----AlaLysProSerGluLys 59
 QY 1061 CCCAGGACGCCACACCCACCTTAGAGAGACAGAGAGAGCTACGAGCAGCAGAG 1120
 :|||||:|||||:|||||: ||| |||
 Db 60 ProArg----- 61
 QY 1121 GTCCTGTGGGGAGGAGCCCTGACATGCCCCGGAGGCTCCACAGAGATGAGAGC 1180
 :|||||:|||||:|||||: ||| |||
 Db 61 ----- 61
 QY 1181 ACACAGAGACCTGTGAGGAGAGCAGCAGAGAGATGGGGAGAGAGAGGATTC 1240
 ::::||||| :|||||:|||||: ||| |||

Db 62 LeuArgGlnIleProSerAlaGluAspLeuGluThrAspGly----- 75
 QY 1241 ATCCAGGTAAAGGAGAGGAGGAGAGAGTGGT-----GCTGAGAGGAGGCGGCATCTG 1294
 :|||||:|||||:|||||: ||| |||
 Db 76 -----GlyGlyProGluGlnValAlaAspAspGlyLeuGluHis 88
 QY 1295 GAGGAGCCGTGCTGCGATATACAAAAGCTGTTCAGATGCCCAACCCGCTGCACCTTTG 1354
 :|||||:|||||:|||||: ||| |||
 Db 89 ArgGluLeuGlnHisGlyGln-----ProGluAlaArgGlyProAlaProLeu 104
 QY 1355 CAGGTGTACCAAGCCGCC----- 1372
 :|||||:|||||:|||||: ||| |||
 Db 105 Gln-----GlnHisProGlnValLeuLeuThrGluGlnArgLeuAlaGlyArgLeu 122
 QY 1373 -----CTCACCCTGGCCACGTGGTCCCGCCACCAACCC 1402
 :|||||:|||||:|||||: ||| |||
 Db 123 ProArgGlySerThrGlyAspThrValLeuLeuProLeuAlaGlnGlyGlnHisArgPro 142
 QY 1403 CTGGGCGCTACCCAAATCTCTCCCTGCTGCGCTGGGGGCGCATGAAGAACCC----- 1453
 :|||||:|||||:|||||: ||| |||
 Db 143 LeuSerArgAlaGlnSerSerProAlaAlaProAlaSerLeuSerAlaProGluProAla 162
 QY 1454 -----CCAGACCAACCCGTCMAACACCTC---TTCACACA 1486
 :|||||:|||||:|||||: ||| |||
 Db 163 SerGlnAlaArgValLeuSerSerSerGluThrProAlaArgThrLeuProPheThrThr 182
 QY 1487 ACTGTGGCTCTAGACACGTCTATCTTAAGACCCAGTGCATGTGGGGACACACACAGCTG 1546
 :|||||:|||||:|||||: ||| |||
 Db 183 GlyLeuIleTyrAspSerValMetLeuLysHisGlnCysSerCysGlyAspAsnSerArg 202
 QY 1547 CACCTGTAGCATGCTGGCGCGATCCAGACGATCTGTCGCGGCTCCAGAGACAGCGCTG 1606
 :|||||:|||||:|||||: ||| |||
 Db 203 HisProGlnHisAlaGlyArgGlnGlnSerIleTrpSerArgGlnGlnArgGlyLeu 222
 QY 1607 CTTCAGAAAGTGGCAGCGGATCCGAGTGCAGAAAGCCACGATGATGATGATCCAGACAGT 1666
 :|||||:|||||:|||||: ||| |||
 Db 223 LeuSerGlnCysGluCysLeuArgGlyArgGlyAlaSerLeuGlnGluGlnInsVal 242
 QY 1667 CACTCTGAATACACACCTCTGTATGGGACCAATCCCTCAACCGGCAGACAGCTAATAC 1726
 :|||||:|||||:|||||: ||| |||
 Db 243 HisSerGlnArgHisValLeuLeuTyrGlyThrAsnProLeuSerArgLeuLysAsnSp 262
 QY 1727 AGCAGAGAGTTCGCTGCGTCCCATCAGCCAGCAAGATGATGCTGCTGCTGGGGGC 1786
 :|||||:|||||:|||||: ||| |||
 Db 263 AsnGlyLysLeuAlaGlyLeuLeuAlaGlnArgMetPheValMetLeuProCysGly 282
 QY 1787 ATCGGGGTGACAGTGAACCGTGTGATGATGATGATGATGATGATGATGATGATGATG 1846
 :|||||:|||||:|||||: ||| |||
 Db 283 ValGlyValAspThrAspThrIleTrpAsnGlnLeuHisSerSerAsnAlaAlaArgTrp 302
 QY 1847 GCAGTGGCGTCTGCTGAGTGCCTTCAAGGTGGCTGCAGAGAGACTCAAGAATGGA 1906
 :|||||:|||||:|||||: ||| |||
 Db 303 AlaAlaGlySerValThrAspLeuAlaPheLysValAlaSerArgGluLeuLysAsnIly 322
 QY 1907 TTGGCATCATCCGCGCCCGCCAGACACACGACCGCGGAATGACACAGCCATGGAGTTCTGC 1966
 :|||||:|||||:|||||: ||| |||
 Db 323 PheAlaValAlaArgProProGlnHisHisAlaAspHisSerThrAlaMetClyPheCys 342
 QY 1967 TTCCTCAACTGTAGCCATCACCGCAAAACTCTACAGCAGAGAGTGAACGTGGGCAAG 2026
 :|||||:|||||:|||||: ||| |||
 Db 343 PhePheAsnSerValAlaIleAlaCysArgGlnLeuGlnGlnInsLysAlaSerLys 362
 QY 2027 GTCCTCATCGTGATGGGAGATTTACCATGTGCATGGCACCACGAGCGTTCATCAAT 2086
 :|||||:|||||:|||||: ||| |||
 Db 363 IleLeuIleValAspTrpAspValHisHisGlyAsnGlyThrGlnThrPheThrGln 382
 QY 2087 GACCCCTGTGCTCTATATCTCTGATGCTGTGACAGAGGAACTTCTTCCAGGC 2146
 :|||||:|||||:|||||: ||| |||
 Db 383 AspProSerValLeuTyrIleSerLeuHisArgHisAspAspGlyAspPheProGly 402
 QY 2147 TCTGGGCTCTGAGAGAGTGTGTGAGACCGAGCGGTGGGTACATGTAGACGTGCA 2206
 :|||||:|||||:|||||: ||| |||
 Db 403 SerGlyAlaValAspGluValGlyAlaSerSerGlyGlnGlyPheAsnValAlaInsValAla 422

[illegible][illegible]


```

Db 230 LeuAlaGlnArgMetPheValMetLeuProCysGlyValGlyValAspThr 249
QY 1808 GTGGGAATGATGCACTCTCCATGCTGTGGCGCATGGCAGTGGCTGCTGGAG 1867
Db 250 ILerPrasnGluLeuHisSerSerAsnAlaIArgTrpAlaAlaGlySerValThAsp 269
QY 1868 CTGGCCCTTCAGAGTGGCTGGAGAGTGGCAAGATTCGCAATTCGCGCCGCCCA 1927
Db 270 LeuAlaPheLysValAlaSerArgGluLeuLysAsnGlyPheAlaValAlaArgProPro 289
QY 1928 GGACACCAAGCCGAGAGATCCACAGCCATGGAGATTCGCTTCCTCAACTGTAGCCATC 1987
Db 290 GlyHisHisAlaAspHisSerThrAlaMetGlyPheCysPhePheAsnSerValAlaIle 309
QY 1988 ACCGCAAACTCCACAGCAAGATTCAGCTGGGCAAGTCCATCCATCGTGGAGAC 2047
Db 310 AlaCysArgGlnLeuGlnGlnSerLysAlaSerLysIleLeuIleValAspTrpAsp 329
QY 2048 ATTCACCATGGCAATGGCAGCCAGCAGCGCTTCATCAATGACCCCTGCTGTACATC 2107
Db 330 ValHisHisGlyAsnGlyThrGlnGlnThrPheTyrgLinsProSerValLeuTyrlle 349
QY 2108 TCTCTGCATGCTATGACAAAGGGAAGTCTTTCACAGCTCTGGGCTCTGAAGAGTT 2167
Db 350 SerLeuHisArgHisAspAspGlyAsnPhePheProGlySerGlyAlaValAspGluVal 369
QY 2168 GGTGGAGACAGCGGCTGGGTTACATGTAAGTGGCATGGCAGACAGAGTGGAGACCC 2227
Db 370 GlyAlaGlySerGlyGlyGlyPheAsnValAsnValAlaTrpAlaGlyGlyLeuAspPro 389
QY 2228 CCCATGGACAGCTGGAGTACCTTCACAGCTTCAGACAGAGTGGATGCCATGCCAC 2287
Db 390 PrometGlyAspProGlyGlyLeuAlaAlaPheArgGlyLeuValMetProIleAlaArg 409
QY 2288 GAGTTTCACCTGATGTGTGTCTAGTCTCCGCGGGTTGATGCTGTGAAGACATCTG 2347
Db 410 GluPheSerProAspLeuValLeuValSerAlaGlyPheAspAlaAlaGlyGlyHisPro 429
QY 2348 TCTCTCTGGTGGCTACTGTCTGTCAGCCGATGTTTGGCCACTGGACAGCAGCTG 2407
Db 430 AlaProLeuGlyGlyTrpHisValSerAlaLysCysPheGlyTyrgMetTrpGlnGluLeu 449
QY 2408 ATGACCCCTGGCAGGCGCGGTGTCTGCTGGCCCTGGAGGAGGCCATGACTTGACCGC 2467
Db 450 MetAsnLeuAlaGlyGlyAlaValAlaValLeuGluGlyGlyHisAspLeuThrAla 469
QY 2468 ATCTGTATGCTCTGAAGCTGTGTCTCGGCTCTGCTCACTGTAAGCTGACCCCTTG 2527
Db 470 IleCysAspAlaSerGluAlaCysValAlaAlaLeuLeuGlyAsnArgValAspProLeu 489
QY 2528 GATGAGCAGCTCTTCGCAAAAGCCCAACATCAAGCAGTGGCGCCATGAGCAAAAGTC 2587
Db 490 SerGluGluGlyTrpLysGlnLysProAsnLeuAsnSerIleArgSerLeuGluAlaVal 509
QY 2588 ATCGATCCAGAGCAAAACCTGGAGCTGTGTGCAGAAAGTTCCGCGCTGTGGCGCG 2647
Db 510 IleArgValHisSerLysTyrgLysMetGlnArgLeuAlaSerCysProAspSer 529
QY 2648 TCCCTGGAGGCGCCAGCAGGTGAGACCGAAGACCCGAAT-GTGAACGCCATGGCC 2706
Db 530 TrpValProArgValProGlyAlaAspLysGluValGlyAlaValAlaThrAlaLeuAla 549
QY 2707 TTGTGTGTGGGGG-----GCCGACAGGCGCCCAAGCTGGGCGGACCGCGGAACACAGC 2760
Db 550 SerLeuSerValGlyIleLeuAlaGluAsp----- 559
QY 2761 CCCAGCGCGCAGAGACCCATGAGCAGAGAG 2793
Db 560 ---ArgProSerGluGlnLeuValGluGlu 569

```

RESULT 15
AAB42845
ID AAB42845 standard; Protein; 541 AA.

```

XX XX AAB42845;
AC AC
XX XX 08-FEB-2001 (first entry)
DT DT
XX XX Human ORFX ORF2609 polypeptide sequence SEQ ID NO:5218.
DE DE
XX XX
KW KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW KW vulnerable; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
KW KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW KW antiviral; antibacterial; antifungal; antihemetic; antithyroid;
KW KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW KW thrombosis; contraceptive.
XX XX
XX OS Homo sapiens.
XX OS
XX PN MO200058473-A2.
XX PN
XX PD 05-OCT-2000.
XX PD
XX PF 31-MAR-2000; 2000WO-0508621.
XX PF
XX PR 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX PA (CURA-) CURAGEN CORP.
XX PA
XX PI Shinkets RA, Leach M;
XX PI
XX DR WPI; 2000-602362/57.
XX DR N-PSDB; AAC77034.
XX PT
XX PT Novel nucleic acids and peptides derived from open reading frame X,
XX PT useful for treating e.g. cancers, proliferative disorders,
XX PT neurodegenerative disorders and cardiovascular disease -
XX PS
PS Claim 11; Page 4402-4403; 5507pp; English.
XX XX
XX AC AACT4446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,
XX CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX CC sequences have activities such as: cytostatic; hepatotropic; vulnerable;
XX CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
XX CC osteopathic; anticonvulsant; antidiabetic; immunosuppressant;
XX CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX CC antidiabetic; hypotensive; dermatological; immunosuppressive;
XX CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
XX CC antithyroid; and antianemic. The sequences can be used for determining
XX CC the presence of or predisposition to, or preventing or treating
XX CC pathological conditions associated with an ORFX-associated disorder. The
XX CC nucleic acids can be used to express ORFX proteins in gene therapy
XX CC vectors. The proteins and nucleic acids may be used to treat cancers,
XX CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX CC graft vs host disease, cardiovascular disease, diabetes mellitus,
XX CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
XX CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX XX
XX SQ Sequence 541 AA;

```

Alignment Scores:

GenCore version 5.1.4 p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 21, 2003, 12:20:42 ; Search time 53.2241 Seconds
(without alignments)
12016.349 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732
Sequence: 1 cttctgcatgcattccgagaaa.....aaatgacattttaagaagaa 1552

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-O/cgn2.1/USFTO.spool/US0502945/runat_14032003_101058_19113/app_query.fast_1.10979
-DB=SPREMBL_21 -QFMT=fastan -SUFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USFR=US0502945 @cgn 1.1.565 @runat_14032003_101058_19113 -NCPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -MAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: SPREMBL_21:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.potent:*
13: sp.virus:*
14: sp.vertebrate:*
15: sp.unclassified:*
16: sp.virus:*
17: sp.bacteriophage:*
18: sp.archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2062	75.5	414	4 060527	060527 homo sapien

2	935	34.2	468	11 09CSM9	09CSM9 mus musculus
3	841	30.8	390	4 09POF1	09POF1 homo sapien
4	521	19.1	379	11 08R3K0	08R3K0 mus musculus
5	343	12.6	314	11 09CYH0	09CYH0 mus musculus
6	272.5	10.0	178	5 076329	076329 dictyostell
7	256.5	9.4	1790	3 007380	007380 saccharomyc
8	254	9.3	2017	5 094992	094992 drosophila
9	254	9.3	2056	5 09K0W8	09K0W8 drosophila
10	254	9.3	2057	5 094987	094987 drosophila
11	253.5	9.3	1931	13 091973	091973 coturnix co
12	253.5	9.3	3259	4 014789	014789 homo sapien
13	252	9.2	1022	11 09WVCO	09WVCO mus musculus
14	251	9.2	876	5 09BMV8	09BMV8 sarcopetes s
15	250.5	9.2	891	13 090MH5	090MH5 gallus gall
16	250.5	9.2	1931	13 0910C5	0910C5 gallus gall
17	250.5	9.2	3616	13 09W6V0	09W6V0 gallus gall
18	249.5	9.1	2007	13 002015	002015 gallus gall
19	249	9.1	1294	5 09VXU1	09VXU1 drosophila
20	249	9.1	1398	5 0960D0	0960D0 drosophila
21	248	9.1	2611	11 091208	091208 mus musculus
22	246.5	9.0	682	5 096720	096720 dermatophag
23	246.5	9.0	1581	4 092614	092614 homo sapien
24	246.5	9.0	1992	13 004834	004834 xenopus lae
25	245	9.0	1583	4 015045	015045 homo sapien
26	244.5	8.9	1964	5 08SMQ7	08SMQ7 loligo peal
27	244	8.9	688	4 096NL6	096NL6 homo sapien
28	244	8.9	764	13 091411	091411 gallus gall
29	243.5	8.9	625	6 002717	002717 bos taurus
30	243.5	8.9	1229	5 09N2J2	09N2J2 aequipecten
31	243.5	8.9	1243	5 09N2J1	09N2J1 aequipecten
32	243.5	8.9	1253	5 09N2J0	09N2J0 aequipecten
33	243.5	8.9	1588	11 09ESK9	09ESK9 mus musculus
34	243.5	8.9	1951	5 017042	017042 aequipecten
35	241.5	8.8	1939	13 09PTY2	09PTY2 gallus gall
36	241.5	8.8	2442	4 060588	060588 homo sapien
37	241.5	8.8	2442	4 09H450	09H450 homo sapien
38	241	8.8	1313	4 075033	075033 homo sapien
39	241	8.8	1937	5 025142	025142 halocynthia
40	241	8.8	1940	5 002456	002456 schistosoma
41	240.5	8.8	1929	13 098706	098706 notiothena
42	240.5	8.8	1941	13 09DGM4	09DGM4 gallus gall
43	240.5	8.8	2029	4 09C014	09C014 homo sapien
44	240.5	8.8	2069	4 09C013	09C013 homo sapien
45	240.5	8.8	2073	4 09C012	09C012 homo sapien

ALIGNMENTS

RESULT 1	ID	Score	Query Match	Length	DB ID	Description
060527	060527	75.5	414	4 060527	060527	060527
AC	060527					
DT	01-AUG-1998 (TREMBlrel. 07, Created)					
DT	01-AUG-1998 (TREMBlrel. 07, Last sequence update)					
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)					
DE	Antigen NY-CO-8 (Fragment).					
GN	NY-CO-8.					
OS	Homo sapiens (human).					
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OC	NCBI_TaxId=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=COLORECTAL CARCINOMA;					
RX	MEDLINE=98272252; PubMed=9610721;					
RA	Scanlan M.J., Chen Y.T., Williamson B., Gure A.O., Stockert E.,					
RA	Gordan J.D., Tureci O., Sahin U., Pfeundschn M., Old L.J.;					
RT	"Characterization of human colon cancer antigens recognized by					
RL	autologous antibodies.";					
DR	Int. J. Cancer 76:652-658(1998).					
DR	EMBL; AF039690; AAC18039.1; -.					
FT	NON_TER					
SO	SEQUENCE					

Alignment Scores:

Pred. No.:	8,58e-124	Length:	414
Score:	2062.00	Matches:	414
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	75.48%	Indels:	0
DB:	4	Gaps:	0

US-09-502-945-1 (1-1552) x 060527 (1-414)

```

OY 1 CTTTCGATGATCCGAGAGAACTTAATCTTATGAGAGAAAGTGTGAATTCAGAA 60
    |||||||
DB 1 LeuLeuAspAlaSerGlnLysLeuLysLeuThrTyrGlnGlnLysCysGlnLysGln 20
OY 61 TCCCAATTAAGCTTTTGGAGAACGCTTACGGAATATCAGAGAACTTGTGAAGTCTT 120
    |||||||
DB 21 SerGlnLeuLysPheLeuLysArgAsnAspLeuAlaGlnTyrGlnArgThrCysGlnAspLeu 40
OY 121 AAGAGCAACTAAGCAATTAAGCAATTTCTTGCGCTGCTTAATACCTTGAACGCTGTGGT 180
    |||||||
DB 41 LysGlnGlnLeuLysHisLysGlnPheLeuLeuAlaAlaAsnThrCysAsnArgValGly 60
OY 181 GGTCTTTGTTGAATGTCTCAGCATGAGTCTTTCCCAACCCATCTAATGTT 240
    |||||||
DB 61 GlyLeuCysLeuLysCysAlaGlnHisGlnAlaValLeuSerGlnThrHisThrAsnVal 80
OY 241 CATATGACACCATGCAAGAGCTGGTTAAGAGAGATGACTGTATGTCGACTAGTT 300
    |||||||
DB 81 HisMetGlnThrIleGlnLysGlnValLysGlnArgAspAspLeuMetSerAlaLeuVal 100
OY 301 TCCGTAAGGAGAGCTTGGCAGATACGACGAGCAAGAGCAAGCTTATGAACAGGTG 360
    |||||||
DB 101 SerValArgSerSerLeuLysAlaSerThrGlnGlnArgGlnAlaSerAlaTyrGlnGlnVal 120
OY 361 AAGCAAGTTTTCGAATATCTGAGAGAGCCATTTTGAAGAAACCAAGCTTATTCGAG 420
    |||||||
DB 121 LysGlnValLeuGlnIleSerGlnGlnAlaAsnPheGlnLysThrLysAlaLeuIleGln 140
OY 421 TGTGACACATGAGAGAGAGCTGAGAGAGCGAGCGAGCTTGAAGAAACTTGA 480
    |||||||
DB 141 CysAspGlnLeuArgLysLeuGlnLysArgGlnAlaGlnArgLeuLysGlnLeuAla 160
OY 481 TCTCAGCAAGAGAGAGCGCATTTGAGAAACATGATGATGAAAAAGAAATTAACGAAAGAA 540
    |||||||
DB 161 SerGlnGlnGlnLysArgAlaIleGlnLysAspMetMetLysGlnLysThrLysGln 180
OY 541 AAGGAGTCAATGAGATCAAGAGATGTTGATCTGTCTCAAGATATGCCAAGCTGAGAGCC 600
    |||||||
DB 181 ArgGlnLysMetGlnLysSerLysMetLeuIleLeuSerGlnAsnIleAlaGlnLeuGlnAla 200
OY 601 CAGGTGAGAAAGGTTTACAAAGAGAAAGATTTTCACTATTATCAATGAGAGAAATTCAA 660
    |||||||
DB 201 GlnValGlnLysValThrLysGlnLysIleSerAlaIleAsnGlnLeuGlnIleGln 220
OY 661 AGCCAGCTGGCTTCGCGGAATGATGTCACAAAGGTGTGGAGAAATGCGCTATCAG 720
    |||||||
DB 221 SerGlnLeuAlaSerArgLysGlnMetAspValThrLysValCysGlyGlnMetArgTyrGln 240
OY 721 CTGATTAAGCAACATGAGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
    |||||||
DB 241 LeuAsnLysThrAsnMetGlnLysAspGlnAlaGlnLysGlnHisArgGlnPheArgGlnAla 260
OY 781 AAAAATAACAGAGGATCTTGAATTAAGATCAGAAATAGAGAAATTCAGATAGAACTG 840
    |||||||
DB 261 LysThrAsnArgAspLeuGlnLysAspGlnGlnLysGlnLysLeuArgLysGlnLeu 280
OY 841 GATGAAGCAACAACTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
    |||||||
DB 281 AspGlnSerLysGlnHisLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 300
OY 901 TGGCTGAGAGCTTAAGCAAGAACTGTGGCGGAATCTGAGCAGCAACTGCACTCCAGAGTCT 960
    |||||||
  
```

```

DB 301 CysLeuArgLeuThrGlnLeuLeuGlnLysSerGlnHisGlnLeuHisLeuThrArgSer 320
OY 961 GAAATAGCTCAACTCAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
    |||||||
DB 321 GlnIleAlaGlnLeuSerGlnGlnLysArgTyrThrTyrAspLysLeuGlnLysLeuGln 340
OY 1021 AGAAGAAATGAGAAATTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
    |||||||
DB 341 ArgArgAsnGlnGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 360
OY 1081 AAGCAAGGCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
    |||||||
DB 361 LysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnLeuValGln 380
OY 1141 CTCTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
    |||||||
DB 381 LeuLeuSerLysGlnAsnGlnLeuLeuGlnArgGlnSerLeuSerGlnLysValAsp 400
OY 1201 CGGCTGCGGAGCCAGTATACCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1242
    |||||||
DB 401 ArgLeuArgThrGlnLeuProSerMetProGlnSerAspCys 414
RESULT 2
OYCSM9 PRELIMINARY: PRT: 468 AA.
AC 09CSM9;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 2700048621Rik protein (Fragment).
GN 2700048621Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamaki I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikolaic I., Pesole G., Queckenbush J.,
RA Schriml L.M., Stambli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyono K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT *Functional annotation of a full-length mouse cDNA collection.*;
RL Nature 409:685-690(2001).
DR EMBL: AK012392; BAB28209.1; -
DR MGD: MGI:1924066; 2700048621Rik.
FT NON-TER 468
SQ SEQUENCE 468 AA; 53462 MW; DABCEA6808DFBD4 CRC64;
  
```

Alignment Scores:

Pred. No.:	6,61e-52	Length:	468
Score:	935.00	Matches:	184
Percent Similarity:	89.54%	Conservative:	30
Best Local Similarity:	76.99%	Mismatches:	25
Query Match:	34.22%	Indels:	0
DB:	11	Gaps:	0

US-09-502-945-1 (1-1552) x 09CSM9 (1-468)

OY	16	GGAAGCGTAAACCTTACCTATGAGGAAAGTGTGAATTAAGGAGATCCCAATTGAAGTT	75
Db	230		24
OY	76	TTGAGGAACGACTTGTAGCTGATATCCAGAGAACCTTTGAAGATCTTAAAGACAATAAG	13
Db	250		26
OY	136	CATTAAGAATTTCTTCTGGCTGCTAATACCTGTGAACCGTGTGGTGCTTTGTTGAAA	19
Db	270		28
OY	196	TTGTGTCAGCATGAAGCGTGTCTTCCCAAAACCCATACCTAATGTCATATGTCAGACCATC	25
Db	290		30
OY	256	GAAGACCTGTTTAAAGAAAGACATGACTTGATGTGTGCACATGTTCCGTAAGAGACAGC	31
Db	310		32
OY	316	TTTGCAAGATACCCACCAAGAGAACACAGTGTATTGAAACAGGTAAACAGTTTGCAA	37
Db	330		34
OY	376	ATATCTGAGGAAGCCCAATTTTGAATAAACCAAGCGTTTAAATCCAGTGTACAGCTTGAG	43
Db	390		36
OY	436	AAGGAGCTGGAGAGGACGCGGAGCGACTTGAAAAAGAACTGTCATCTCAGACAGAGAA	49
Db	370		38
OY	496	AGGCCCATTTGGAAGACATGATGAAAAAGAAATACGAAAGAAAGGAGTACATGGGA	55
Db	390		40
OY	556	TCAAGAGATGTGATCTTGTCTCAGAAATATTTGCCCACTGAGAGCCACAGTGGAAAGCTT	61
Db	410		42
OY	616	ACAAAGAAAGATTTCAAGCTATTATATCACTGAGAGAAATTCAAAGCCACCTGGCTTCT	67
Db	430		44
OY	676	CGGGAATAGGATGTGCACAAAGGTGTGTGGAGAAATGCCCTATCAGCTAATTAACCC	732
Db	450		468

RESULT 3

O9Pop1 PRELIMINARY: PRI: 393 AA.

AC O9Pop1. TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE HSPC085 (Fragment).

OS Homo sapiens (Human).

OC Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID:9606;

PN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=CORD BLOOD;

RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,

RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;

RT "Human partial CDS cloned from cd34+ stem cells.";

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF161348; AAF28908.1; -.

FT NON_TER 393

SO SEQUENCE 393 AA; 44926 MW; FFFD0D7090F39666 CRC64;

Pred. No.:	6,586-46	Length:	393
Score:	841.00	Matches:	168
Percent Similarity:	100.008	Conservative:	0
Best Local Similarity:	100.008	Mismatches:	0
Query Match:	30.78%	Indels:	0
DB:	4	Gaps:	0

US-09-502-945-1 (1-1552) x Q9P0F1 (1-393)

QY	16	GAGAAAGCTAAACCTTACTATATGAGAAAGAGTGTGAAATGAGAAATCCCAATTTGAAGTTT	75
Db	226	GLIUTSLEULysLEUthrTYrctGtGULysCYSGLIIEGtGtUserGtInLEUysSPE	24
QY	76	TTGAGGAACGACTTAgCTGAATATGACAGAACTTGAGATCTTAAAGCAACTAAAG	13
Db	246	LeuATgAsnAsPLeuAlaGLIUrTYrGlnATgThrCYSGLIAsPLeuysGLIInLEUys	26
QY	136	CATTAAGAATTTCTTCTGGCTGCTAACTACTTGTAACCGTGGTGtGCTTTGTTGAAA	19
Db	266	HISLysGLIUPheLEUleuAlaIAsnThrCYSAsnATgValGLIyGLIeUysLEUys	28
QY	196	TTGGCTGACATGAAGACTTCTTCCTCCCAAAACCCATACATATGTCATGACAGACATC	25
Db	286	CysAlaGLIInHISGLIuAlaValLEUserGtInThrHISThAsnValHISMetGtInThrILE	30
QY	256	GAAAGACTGGTTAAAGAAAGAGATGACTTGATGTCTGTGCACTGTTCCCTTAAGAGACAC	31
Db	306	GLIATrGLIeValLYsGLIuTrgAsPAsPLeuMetSerAlaLeuValSerAlaTrgSerSer	32
QY	316	TTTGGCAGATAGCCAGCAAAAGACAGACAGTGTCTTATGAACAGGTGAACAACATTTTGGCA	37
Db	326	LeuAlaAsPThrGLIInGLIuATrGLIuAlaSerAlaTYrGLIuGLIInValLYsGLIInValLEUGln	34
QY	376	ATATCTGGAGGAGCCAAATTTTAAAAAACCAAGCGCTTATATCCAGTGCACAGTTGAGG	43
Db	346	IIIESeGLIuGLIuAlaAsnHeGLIuysThrLYsAlaLeuIIIEGLInCysAsPInLEuATg	36
QY	436	AAGGAGCTGTGAAGCGACGCGAGGACACTTGGAAAAAAGACTTCATCTCAGCAAGAGAAA	49
Db	366	LYsGLIuLEuGLIuATrGLIInAlaGLIuATrGLIeUgLYsGLIuLEuAlaSerGLInGLInLYs	38
QY	496	AGGCCCATTTGACAAAGACATGATG 519	
Db	386	ATGAlaIIIEGLIuLYsAsPMetMet 393	
RESULT 4			
Q8R3K0			
AC	Q8R3K0	PRELIMINARY:	PRT: 379 AA.
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Similar to RIKEN cDNA 2700048621 gene.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Strausberg R.;		
RU	Submitted (MAR-2002) to the EMBL/GenBank/DBDJ databases.		
DR	EMBL, BC025138; AAR25138.1, ..		
SO	SEQUENCE 379 AA: 42775 MW: B20D32A9E193F53C CRC64:		

Alignment Scores:

Pred. No.:	1,7e-25	Length:	521.00
Score:	521.00	Matches:	1022
Percent Similarity:	88.55%	Conservative:	14
Best Local Similarity:	77.86%	Mismatches:	15
Query Match:	19.07%	Indels:	0
DB:	11	Gaps:	0

Alignment Scores:

US-09-502-945-1 (1-1552) x Q8R3K0 (1-379)

```

OY 16 GAGAACCTAAACTTACTTATGAGAAAGTGTGAATTTGAGGATCCCAATGAAGTTT 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 192 GUATGLeuLysLeuThrTyrglAlaLysThrAspLeuLysGlnSerGlnLeuMetLeu 211
OY 76 TTGAGGACGACTTACCTGAATATGACAGAACTTGTGAAGATCTTAAAGACAACTAAAG 135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 212 LeuATGLeuLysLeuThrTyrglAlaLysThrCysGlnAspLeuLysGlnATGLeuLys 231
OY 136 CATTAAGAAATTTCTGTGCTGCTATACTTGAACCGTTGGTGGTGGTCTTTGTTGAA 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 232 HsLysGlnSerLeuLeuAlaSerAlaSerSerArgValGlyLysLeuLysCysLeuLys 251
OY 196 TGTGCTCAGCATGAGCTGTCTTCCCAACCCATACATTAATGTTATGACAGCAATC 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 252 CysAlaGlnHsGlnAlaValLeuSerGlnThrHsSerAsnValHsIleGlnThrIle 271
OY 256 GAAAGACTGTTAAAGAAAGACATGCTGATGCTGACCTAGTTCCGTAAGACAGC 315
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 272 GUATGLeuThrLysGlnArgAspAspLeuMetSerValLeuValSerValArgSerSer 291
OY 316 TTGGCAGATACGACGACAAAGACAGACAGCTTATGACAGGTGAACAAAGTTTGCAG 375
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 292 LeuAlaGlnHsGlnAlaLysArgGlnThrSerAlaTyrglGlnValLysHsAlaValGln 311
OY 376 ATATCTGAGGAGCAACCAATTTGAAAAACCAAG 408
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 312 MetThrGlnGlnAlaLysAsnHsGlnLysThrLys 322

```

RESULT 5

O9CYHO PRELIMINARY: PRT: 314 AA.

```

AC O9CYHO;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE 5730470G24RIK.
GN Mus musculus (Mouse).
OS Mus musculus; Chordata; Vertebrata; Euteleostomi;
OC Eukaryota; Metazoa; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikolaic I., Pesole G., Quackenbush J.,
RA Schiml L.M., Staubl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK017691; BAB30877.1;
DR MGD; MGI:1917837; 5730470G24RIK.
SQ SEQUENCE 314 AA: 35730 MW; B41F3B08B8E418F CRC64;

```

Alignment Scores:

Pred. No.: 3 87e-14 Length: 314
Score: 343.00 Matches: 66

Percent Similarity: 88.24% Conservative: 9
Best Local Similarity: 77.65% Mismatches: 10
Query Match: 12.55% Indels: 0
DB: 11 Gaps: 0

US-09-502-945-1 (1-1552) x O9CYHO (1-314)

```

OY 16 GAGAACCTAAACTTACTTATGAGAAAGTGTGAATTTGAGGATCCCAATGAAGTTT 75
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 230 GUATGLeuLysLeuThrTyrglAlaLysThrAspLeuLysGlnSerGlnLeuMetLeu 249
OY 76 TTGAGGACGACTTACCTGAATATGACAGAACTTGTGAAGATCTTAAAGACAACTAAAG 135
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 LeuATGLeuLysLeuThrTyrglAlaLysThrCysGlnAspLeuLysGlnATGLeuLys 269
OY 136 CATTAAGAAATTTCTGTGCTGCTATACTTGAACCGTTGGTGGTGGTCTTTGTTGAA 195
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 270 HsLysGlnSerLeuLeuAlaSerAlaSerSerArgValGlyLysLeuLysCysLeuLys 289
OY 196 TGTGCTCAGCATGAGCTGTCTTCCCAACCCATACATTAATGTTATGACAGCAATC 255
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 CysAlaGlnHsGlnAlaValLeuSerGlnThrHsSerAsnValHsIleGlnThrIle 309
OY 256 GAAAGACTGTTAAAGAAAGACATGCTGATGCTGACCTAGTTCCGTAAGACAGC 315
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 GUATGLeuThrLys 314

```

RESULT 6

O76329 PRELIMINARY: PRT: 1738 AA.

```

AC O76329;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Interactin.
GN ABPD.
OS Dictyostellium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliales; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98365468; PubMed=9700162;
RA Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;
RT "Interactin, an actin-binding protein of the alpha-actinin superfamily
RT in Dictyostellium discoideum, is developmentally and cAMP-regulated and
RT associates with intracellular membrane compartments.";
RL J. Cell Biol. 142:735-750(1998).
DR EMBL; AF057019; AAC34582.1;
DR HSP, P46939; IQAG.
DR InterPro: IPR001589; Actbind_actnin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR001990; Graplin.
DR InterPro: IPR001451; Hexapep_transf.
DR InterPro: IPR002017; Spectrin.
DR pfam: PF00307; CH_2.
DR SMART; SM00033; CH_2.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; GRANIN_1; UNKNOWN_1.
DR PROSITE; PS00422; GRANINS_1; UNKNOWN_1.
DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; UNKNOWN_1.
SQ SEQUENCE 1738 AA: 204427 MW; 577A99D2EC79AF5C CRC64;

```

Alignment Scores:

Pred. No.: 1 15e-09 Length: 1738
Score: 272.50 Matches: 106
Percent Similarity: 42.16% Conservative: 109
Best Local Similarity: 20.78% Mismatches: 162
Query Match: 9.97% Indels: 133
DB: 5 Gaps: 17

US-09-502-945-1 (1-1552) x O76329 (1-1738)

```

QY 13 TCCGAGAACTAAACTTACTTATGAGGAAAAGTGTGAATTTGAGAAATGCCAATTGAG 72
Db 844 AsnGlnLysAsnGlnLysAspHisGlnAspGlnLeuGlnLeuGlnLysGlnLeuLys 863
QY 73 TTTTGGAGAACGACTTATGCTGAATATCAGAACTTGTGAAGATCTTAAAGACCACTA 132
Db 864 GlnLeuGlnGlnGlnLysAspGlnLeuAsnGlnLysAsnGlnSerIleGlnAsnGlnLeu 883
QY 133 AAGCATTAAGAAATTTCTGCTGCTGCTTACTTGTAACTGTAACCTGTGGTCTTGTGTTG 192
Db 884 AsnGlnGlnAsnLeuLysAsnLysGlnAsn -----Leu 894
QY 193 AAATGCTCAGCATGAGCTGTTCTTCCCAAAACCATTAATGTTCAATGACAGACC 252
Db 895 AsnGlnLysGlnGlnLysLeuLeuLysLeuGlnAsnGlnLeuAsnGlnGlnLys 914
QY 253 ATCGAA -----AGCTGCTTAA 270
Db 915 IleGlnPheAspGlnGlnLysPheSerLysGlnAsnSerIleAsnIleGlnLeuValAsn 934
QY 271 GAAAGAGATGACTGTGATGTCTGACTAGTTCCGTAAGAGACAGCTGGCAGATACGACG 330
Db 935 GlnLysAsnGln -----LysLeuIleGlnLeuGln 944
QY 331 CAAAGAGAACGCAAGTCTTATGAACAGGTGAACAAAGTTTTCAAATATCTGAGAAAGCC 390
Db 945 Gln -----AspTyrAspGlnLeuLysGlnGlnAsnAspGlnLys 960
QY 391 AAT -----TTGAAAAAACCAAGCTTTAATCCAGTGTGACCACTTGAGCAG 438
Db 961 AspLysAsnAspLeuIleGlnLysGlnAsn -----GlnLeuLysSerIleGlnAsn 977
QY 439 GAGCTGGAGAGCAGCGGCGGACTGTAAGAAAGAACTTCATCAGCA -----GAGAA 495
Db 978 GlnLeuAsnGlnLysIleGlnLysAsnGlnSerAspHisLysGlnGlnLysGln 997
QY 496 AGGCGCATTTGAGAAACATGATGAAAAAGAA ----- 528
Db 998 GlnSerIleGlnAsnAspLeuIleGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGln 1017
QY 529 ATAAAGAAAGAAAGGATGATGATGATCAAAAGATGTTGATCTTGTCTCAGAAATATGCC 588
Db 1018 LeuAsnGlnGlnGlnGlnGlnGlnGlnGlnGln -----LeuSerGlnLysAspGln 1034
QY 589 CAACTGGAGGCCAGGTGGAAG ----- 612
Db 1035 GlnLeuAsnGlnLysIleGlnLysAsnGlnPheAspGlnLysGlnGlnGlnLysGln 1054
QY 613 -----GTACAAAGAAAGAAATTTCAAGCTTATATCACTGAGGAAATTCAAAGCCAG 666
Db 1055 GlnSerIleGlnAsnAspLeuPheGlnLysGlnAsnGlnIleGlnGlnGlnGlnGln 1074
QY 667 CTGCTCTCTCGGAAATGATGTCAAAGGTGTGTGAGAAATGCCATACCTGAT 726
Db 1075 LeuAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnLeuSerGlnLysAspGlnGlnLeuAsn 1094
QY 727 AAAACCAACATGAGAGATGAGCAGAAAGAAAGCAGCAGACAGCTCAGCAAAAAC 786
Db 1095 Gln -----IleGlnLysAsnGlnSerAspGlnLysGlnGlnGlnGlnGlnGlnGln 1113
QY 787 ---AACAGGATCTTGAATTAAGATCAGGAATAGAAATGAGAAATGAGACTGGAT 843
Db 1114 IleGlnAsnAspLeuIleGlnLysGlnAsnGlnIleGlnGlnGlnGlnGlnGlnGln 1133
QY 844 GAAAGCAAAACA ----- 855
Db 1134 GlnGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1153
QY 856 -----CACTGGAACAGGAGCAGCAG 876
Db 1154 LysGlnLeuLysGlnGlnGlnSerAspLeuLysLeuAsnAspGlnLysGlnGlnGln 1173
QY 877 AAGCAGCCCTGCGCAGAGAGAGAGTGC -----CTGAGACTTAAACA ----- 915

```

```

Db 1174 AspLysGlnLeuGlnAspLysGlnIleGlnPheAspGlnLeuGlnLeuPheAsnGln 1193
QY 916 -----CACTGCTGGCGCAATCTGACACCAACTG 945
Db 1194 PheLysAsnAspLysAspSerGlnPheIleGlnLeuGlnAsnAspGlnLysGlnGlnLeu 1213
QY 946 CACCTCACAGATCTGAAATAGCTCAACTCAGTCAAGAAAAAGTATATCATAT ----- 999
Db 1214 GlnSerIleGlnGlnAspLeuAsnGlnLeuLysGlnGlnGlnGlnGlnGlnGlnGln 1233
QY 1000 -----GATAAATGGAAGTATACAGAGAAAGAAATGAAATGAGAGAA 1044
Db 1234 LeuSerGlnLysAspGlnLysLeuGlnGlnSerIleGlnPheGlnAsnGlnGlnGlnLys 1253
QY 1045 CAGTGTCTCCAGCATGGAGAGTACATGACATGAAAGCAAAAGCAAGGACCTGAT 1104
Db 1254 GlnLeuSerGlnLysAspGlnLysLeuGlnGlnSerIleGlnGlnAsnLeuAsnGlnLeuAsn 1273
QY 1105 AAGCAGACGCGCCAGCCAGCCAGCAG -----CTGCTCAGCTC 1143
Db 1274 AspLysAsnGlnGlnLysValLysGlnPheSerLysAspGlnLysLeuGlnSerIle 1293
QY 1144 CTCAGCAAGCAGAACCAAGCTTCTCTGAGAGGAGAGCCGTGTGGAAGAGGTGACCGG 1203
Db 1294 GlnGlnAspLeuAsnGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1313
QY 1204 CTGCGGACCCAGTTACCCAGCATGCCACAA 1233
Db 1314 LysAspGlnLysLeuGlnSerIleGlnGln 1323

RESULT 7
ID 007380 PRELIMINARY; PRT: 1790 AA.
AC 007380; P89892;
DT 01-NOV-1996 (TREMBLrel, 01, Created)
DT 01-NOV-1996 (TREMBLrel, 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel, 21, Last annotation update)
DE Hypothetical 206.5 kDa protein YDLO58W.
GN USO1 OR YDLO58W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId:4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Bloeker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 274106; CAA98621.1; -.
DR EMBL; 274105; CAA98620.1; -.
DR SCD; S0002216; USO1.
DR InterPro; IPR002017; Spectrin.
SQ SEQUENCE 1790 AA; 206450 MW; 90062544F55A52EE CRC64;

Alignment Scores:
Pred. No.: 1, 21e-08 Length: 1790
Score: 256.50 Matches: 86
Percent Similarity: 43.33% Conservative: 99
Best Local Similarity: 20.14% Mismatches: 169
Query Match: 9.39% Indels: 73
DB: Gaps: 11

US-09-502-945-1 (1-1552) x Q07380 (1-1790)
QY 64 CAATTTGAATTTTGGAGAACGACTTATGCTGAATATACAGAACTTGTGAAGATCTT--- 120
Db 963 LysLeuLysSerLeuAlaAsnAsnLysLysAspMetGlnIleGlnGlnGlnGlnGlnGln 982
QY 121 -----AAAGAGCAACCAAGCAATTAAGAAATTTCTGCTGCTGCTATATCTGTAACCGT 174
Db 983 LysAlaValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1002
QY 175 GTTGTGTCTTTGTTGAATGTGCTCAGCATGAAAGCTGTTCTTCCAAACCATACT 234

```

D	b	1003	Il	a	s	p	s	e	r	-----	:::	Me	s	e	r	G	i	n	g	l	u	s	e	l	u	1011																											
O	y	235	A	A	T	G	T	C	A	T	G	C	A	G	-----	-----	A	C	A	T	G	A	A	A	G	C	T	G	T	T	A	A	G	A	A	G	A	276															
				:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:																	
D	b	1012	A	s	n	p	e	c	h	i	n	e	g	i	u	a	r	g	i	s	e	r	I	e	g	l	u	y	s	a	n	I	e	g	l	u	l	e	n	l	y	s	t	r	i	l	e	1031					
O	y	277	G	A	T	C	A	T	G	A	C	A	G	T	T	O	C	G	T	T	A	G	A	G	C	G	T	T	G	G	A	T	A	C	G	A	G	A	A	G	A	336											
				:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:											
D	b	1032	S	e	r	a	s	p	e	u	l	e	u	g	i	n	t	h	r	L	y	s	g	i	u	l	i	e	I	e	S	e	r	L	y	s	e	r	a	s	p	e	u	1051									
O	y	337	G	A	A	C	A	G	T	G	T	T	A	G	A	C	A	G	T	G	A	A	A	A	G	T	T	T	C	A	A	T	A	T	C	T	G	A	G	A	C	A	A	T	T	396							
				:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:											
D	b	1052	T	y	r	i	u	s	e	r	G	i	n	I	e	s	e	r	L	e	u	l	e	u	y	s	G	i	u	l	y	s	e	u	g	i	n	t	h	r	A	l	a	s	n	p	1071						
O	y	397	G	A	A	A	A	A	C	A	A	G	C	T	T	T	A	A	T	C	A	G	T	G	A	C	C	A	C	T	T	A	G	A	A	G	A	C	T	G	A	G	C	A	453								
				:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:											
D	b	1072	G	i	u	s	n	v	a	l	a	s	n	l	y	s	I	e	s	e	r	G	i	u	l	e	u	t	h	r	T	h	r	A	r	g	i	u	l	e	u	l	e	u	1091								
			-----	:	:	:	:	:	:	:	:	-----	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:											
O	y	454	-----	:	:	:	:	:	:	:	:	-----	G	C	G	A	G	G	A	C	T	T	465																														
D	b	1092	A	l	a	a	r	T	r	y	s	a	n	l	e	u	y	s	a	n	g	i	u	l	e	u	g	i	n	t	h	r	L	y	s	e	u	g	i	n	t	h	r	S	e	r	i	u	l	y	s	a	1111
O	y	466	G	A	A	A	A	G	A	C	T	T	G	C	A	T	C	A	G	A	G	A	A	A	G	G	C	C	A	T	T	A	G	A	A	A	A	G	A	C	A	T	G	A	A	A	G	525					
				:	:	:	:	:	:	:	:		:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:											
D	b	1112	L	e	u	y	s	G	i	u	a	l	L	y	s	G	i	u	a	n	g	i	u	l	i	n	I	e	u	l	y	s	e	u	g	i																	

Db	1335	ArgleuylsYlstrHrserSerglu-----GlualgYlsasmlaaglucInleu	1351
Qy	1198	GACCGGCTGGCGACCGACGTTA	1218
Db	1352	GlulYlsleuYlsasngluille	1358
RESULT 8			
	094992	PRELIMINARY;	PRT; 2017 AA.
ID	Q94992		
AC	Q94992; 024138;		
DT	01-FEB-1997 (TREMBLrel_02, Created)		
DT	01-FEB-1997 (TREMBLrel_02, last sequence update)		
DE	01-JUN-2002 (TREMBLrel_21, last annotation update)		
DE	Nonmuscle myosin-II heavy chain.		
GN	ZIP OR CG15792.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachyocera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_Taxid=7227;		
RA	[[1]]		
RE	SEQUENCE FROM N.A.		
RX	MEDLINE=6144835; PubMed=8568878;		
RA	Mansfield S.G., al-Shirawi D.Y., Ketchum A.S., Newbern E.C.,		
RA	Richard D.P.;		
RT	"Molecular organization and alternative splicing in zipper, the gene		
RT	that encodes the Drosophila non-muscle myosin II heavy chain.";		
RL	J. Mol. Biol. 255:98-109(1996).		
DR	EMBL; U35816; AAB09048.1; -.		
DR	EMBL; U35816; AAB09051.1; -.		
DR	HSSP; P10587; 1BR2.		
DR	Flybase; FBgn0005634; zip.		
DR	InterPro; IPR000048; IQ_region.		
DR	InterPro; IPR001609; myosin_head.		
DR	InterPro; IPR004009; myosin_N.		
DR	InterPro; IPR002928; myosin_tail.		
DR	Pfam; PF00612; IQ; 1.		
DR	Pfam; PF00063; myosin_head; 1.		
DR	Pfam; PF02736; Myosin_N; 1.		
DR	Pfam; PF01576; Myosin_tail; 1.		
DR	PRINTS; PR00193; MYOSINHEAVY.		
DR	ProDom; PD000355; myosin_head; 1.		
DR	SMART; SM00015; IQ; 1.		
DR	SMART; SM00242; MYSC; 1.		
DR	PROSITE; PSS0096; IQ; 1.		
DR	SEQUENCE 2017 AA; 232089 MW; 8C76FEFZEBD02EBE CRC64;		
Alignment Scores:			
	Pred. No.:	1.73e-08	Length:
	Score:	254.00	Matches:
	Percent Similarity:	43.15%	Conservative:
	Best local Similarity:	22.49%	Mismatches:
	Query Match:	9.30%	Indels:
	Db:	5	Gaps:
			18
US-09-502-945-1 (1-1552) x Q94992 (1-2017)			
Qy	4	CTGATGTCATCCCGAGCGTAAACTTACTTATGAGAAAGTGGAATTGAGATCC	63
Db	1143	IlleapglucInseralatrHrLysalatrHralaGlInLysalagInhrGluInleuInser	1162
Qy	64	CAATTGAAGTTTGTGAGAGCACTA-----GCTGAATATCAG	102
Db	1163	GlnleuAlaGluIleGlnGlnAspLeuGlnAlaGluLysAlaAlaLargAlaLysAlaGlu	1182
Qy	103	AGAACTTGGAAGATCTTAAAGCACTPAAG-----CATAAAGAAATTCTTGCGCT	156
Db	1183	LysValaLrGArGAspLeuSerGInGluLeuGlnAlaLeuLysAsnGlnleuLeuAspEr	1202
Qy	157	GCTAAATCTTGTAACCTGTGGTGGCTTTGTTGCAATGCGTCAG-----	204
Db	1203	LeuAspHrThrAlaAlaGlnGlnGlnleuLeuAspSerLysArgGluGlnleuAlaThr	1222

QY 205 -----CATGAGCTGTTCTT----- 219
 Db 1223 LeuLysSerLeuGluGluThrValAsnHisGluGlyValLeuAlaAspMetArg 1242
 QY 220 -----TCCCAAAACCCATCTAATGTCATATGACACCATTGAAAGACTGTTAA 270
 Db 1243 HisLysHisSerGlnLeuAsnSerIleAsnAspGln-----LeuGluAsnLeuArgLys 1261
 QY 271 -----GAAAGAGATGACTGTATG 288
 Db 1262 AlaLysThrValLeuGluLysAlaLysGlyThrLeuGluAlaGluAsnAlaAspLeuAla 1281
 QY 289 TCTGCACACTGTTCCGTAAGAGACAGCTTGCGATACGACAAAGAGAGACAGAGCTCT 348
 Db 1282 ThrGluLeuArgSerValAsnSerSerArgGlnGluAsnAspArgArgLysGlnAla 1301
 QY 349 TATGCAACGTTGAAACAGTTTGCATATCTGAGAGCCCAATTTGAAAAACCAAG 408
 Db 1302 GluSerGlnIleAlaGlu-----LeuGlnVal-----LysLeuAlaGluIleGluArgLys 1319
 QY 409 GCTTAAATCCAG-----TGTGACAGTTGAGGAAGAGCGTGGAGGACGCGAGCGA 462
 Db 1320 SerGluLeuGlnGluLysCysThrLysLeuGlnGlnGluAlaAsnIleThrAsnGln 1339
 QY 463 CTTGAAAA-----GAAGCTTCATCTCAGCAAGAGAAA 495
 Db 1340 LeuGluGluAlaGluLeuLysAlaSerAlaValLysSerAlaSerAsnMetGluSer 1359
 QY 496 AGGCCATTTGGAAGACATGATGAAAAAGAAATACGAAAGAAAGGAGCTACATGGGA 555
 Db 1360 GlnLeuThrGluAlaGlnGlnLeuGlnGluGlnThrArgGlnLys----- 1375
 QY 556 TCAAGAGTGTGATCTGTCAGAAATATGCCCACTGAGGCGCCAGCTGGAAAAAGTT 615
 Db 1376 -----LeuGlyLeuSerSerLysLeuArgGlnIleGluSerGlnLysGlnAlaLeu 1392
 QY 616 ACAAGAGAAAGATTTTACGATTAATCACTGAGAGAAATTCAGAGCCAGCTGGCTCT 675
 Db 1393 GlnGlu-----GlnLeuGlnGluAsnAspArgGluAlaLysArgAsn 1405
 QY 676 CGGGAATATGATGTCACAAAGGTGTGTGAGAAATGCCCTTACGCTGATTAACCAAC 735
 Db 1406 TyrGluArgLysLeuAlaGlnValThrThrGlnMetGln-----GluIleLysLysAla 1424
 QY 736 ATGGAAGAGATGAT 795
 Db 1425 GluGluAspAlaAspLeuAlaLysGlnLeuGlnGlnGlnLysLysArgLeuAsnLysAsp 1444
 QY 796 CTTGAATTAAGATCAGAGAAATAGAGAAATTCAGATAGAA-----CTGGATGAA 846
 Db 1445 IleGluAlaLeuGlnArgGlnValLysGlnLeuIleAlaGlnAsnAspArgLeuAspLys 1464
 QY 847 AGCAACACACCTTGGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
 Db 1465 SerLysLysLysIleGlnSerGlnLeuGlnAspAlaThrIleGluLeuGluAlaGlnArg 1484
 QY 886 -----CTGGCCAGAGAG 897
 Db 1485 ThrLysValLeuGlnLeuGlnLysLysLysAsnPheAspLysIleLeuAlaGlnGlu 1504
 QY 898 GAGTGCCGTG-----AGACTAAGAGAACTGCTGGCGAATCTGAGACCAACCTGCACCTC 951
 Db 1505 LysAlaIleSerGlnGlnIleAlaGlnGlnArgAspThrAlaGlnArgGlnAlaArgLys 1524
 QY 952 ACCAGATCTGAATATAGCTCACTCAGTCACAAAAAGAGTATACATATGATTAATGGGA 1011
 Db 1525 LysGluThrLysValLeuSerValSerArgLysLeuAspGluAlaPheAspLysIleGlu 1544
 QY 1012 AAGTTACAGAGAAATGAGAAATGGAGAGACAGTGTGTCAGCATGGAGAGTATCAT 1071
 Db 1545 AspLeuGluAsnLysArgLysThrLeuGlnAsnGlnLeuAspAspLeuAlaAsnThrGln 1564
 QY 1072 GAGACGATGAAGCAAGGCTTAAGCAGCTGATTAAGCACAGCCAGGCCACAGCCAGCAG 1131

Db 1565 GlyThrAlaAspLysAsnValHisGlnLeuGlnLysAlaLysArgAlaLeuGluSerGln 1584
 QY 1132 CTGGTGCAGCTTCCTCAGACAGACAGACAGCTTCTCTGGAGAGAGAGAGAGAGAG 1191
 Db 1585 LeuAlaGlu-----LeuLysAlaGlnAsnGlnGlnLeuGlnGlnAspAspLeuGlnLeuThrGln 1603
 QY 1192 GAGGTGACCGCGCTGGAGCCAGCTTA 1218
 Db 1604 AspAla-----LysLeuArgLeuGluVal 1611
 RESULT 9
 ID 09M08 PRELIMINARY; PRT: 2056 AA.
 AC 09M08:
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE ZIP protein.
 GN ZIP OR CG15792.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup I.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Rehner K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan G., Zhao X., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003465; AAF47311.1;
 DR HSSP: P10587; IBR2.
 DR FlyBase; FBgn0005634; zip.
 DR InterPro; IPR000048; IQ_region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR004009; myosin_N.

[illegible]

DE GIANTIN (GCB372) (MACROGOLGIN) (Golgi autoantigen, golgin subfamily B,
 1).
 GN GOLGI.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9418728; PubMed=7511208;
 RA Seelig H.P., Schrantz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane
 RT protein (giantin) [retracted in Mol Cell Biol 1995 Jan;15(1):591].";
 RL Mol. Cell. Biol. 14:2564-2576(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94257116; PubMed=8198703;
 RA Seelig H.P., Schrantz P., Schroeter H., Wiemann C., Griffiths G.,
 RA Renz M.;
 RT "Macrogolgin--a new 376 kD Golgi complex outer membrane protein as
 RT target of antibodies in patients with rheumatic diseases and HIV
 RT infections.";
 RL J. Autoimmun. 7:67-91(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95100974; PubMed=7802676;
 RA Sonba M., Misumi Y., Fujiwara T., Nishioka M., Iehara Y.;
 RT "Molecular cloning and sequence analysis of a human 372-kDa protein
 RT localized in the Golgi complex.";
 RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).
 CC - FUNCTION: MAY PARTICIPATE IN FORMING INTERCISTERNAL CROSS-BRIDGES
 CC OF THE GOLGI COMPLEX.
 CC - SUBUNIT: DISULFIDE-LINKED HOMODIMER.
 CC - SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.
 CC - ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS MAY ARISE BY ALTERNATIVE
 CC SPLICING.
 CC - DISEASE: ANTIGEN IN CHRONIC RHEUMATOID ARTHRITIS AND IN THE
 CC AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME.
 DR EMBL: D25542; CAAS3052.1; -;
 DR EMBL: D25542; BAA05025.1; -;
 DR MIM: 602500; -;
 KW Golgi stack; Antigen; Coiled coil; Transmembrane;
 KM Alternative splicing.
 FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 3236 3256 POTENTIAL.
 FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).
 FT DOMAIN 48 110 COILED COIL (POTENTIAL).
 FT DOMAIN 127 223 COILED COIL (POTENTIAL).
 FT DOMAIN 238 448 COILED COIL (POTENTIAL).
 FT DOMAIN 460 526 COILED COIL (POTENTIAL).
 FT DOMAIN 545 593 COILED COIL (POTENTIAL).
 FT DOMAIN 677 956 COILED COIL (POTENTIAL).
 FT DOMAIN 969 1028 COILED COIL (POTENTIAL).
 FT DOMAIN 1062 1128 COILED COIL (POTENTIAL).
 FT DOMAIN 1154 1245 COILED COIL (POTENTIAL).
 FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).
 FT DOMAIN 1828 2781 COILED COIL (POTENTIAL).
 FT DOMAIN 2797 2857 COILED COIL (POTENTIAL).
 FT DOMAIN 2872 2993 COILED COIL (POTENTIAL).
 FT DOMAIN 3026 3102 COILED COIL (POTENTIAL).
 FT DOMAIN 3133 3185 COILED COIL (POTENTIAL).
 FT DOMAIN 2420 2423 POLY-GLU.
 FT DOMAIN 2993 2996 POLY-SER.
 FT VARSPLIC 1 39 MISSING (IN REF. 3).
 FT VARSPLIC 215 215 A -> AQLSSM (IN REF. 3).
 FT CONFLICT 1765 1765 D -> G (IN REF. 3).
 FT CONFLICT 2950 2950 H -> D (IN REF. 3).
 SO SEQUENCE 3259 AA; 376075 MW; 60376A20DBA178BD CRC64;

Alignment Scores: 1.84e-08 Length: 3259
 Pred. No.: 253.50 Matches: 119
 Score:

Percent Similarity: 40.33% Conservative: 104
 Best Local Similarity: 21.52% Mismatches: 179
 Query Match: 9.28% Indels: 151
 DB: 4 Gaps: 24
 US-09-502-945-1 (1-1552) x Q14789 (1-3259)
 QY 1 CTTCGGATGTCAGCCGAGAAAGCTTAACCTTACTTAT-----GAG 39
 DB 2204 VALLIENSPGIALALysLysTTPGLVArgLysPheSerAspAlaIleGlnSerLysGlu 2223
 QY 40 GAAAGCTGTGAATTTGAGAAATCCCAATTGAAGTTTGTAGACAGACTTACGTGAATAT 99
 DB 2224 GLUcIUleArgLysLysLysAspAsnGlySerValIleLysAspGlnIleArgGlnMet 2243
 QY 100 CAGAGAACTTGTGAAGCTTTAA-----GAGCACTAAGACAT----- 138
 DB 2244 SerIleHisMetGluGluLysIleAsnIleSerArgLysGlnIleAspLysGlnIle 2263
 QY 139 -----AAGAAATTTCTGTGGCTGTATACTGTAAACCGTGGT 180
 DB 2264 TTPcIUserLysAlaGlnThrGluValGlnIleGlnIleValCysAspThrLysGln 2283
 QY 181 GGTCTTTGTGTAATGTGTCTCAGCAATGAGCTGTCTTTCCCAACCCATATATGTT 240
 DB 2284 GLY-----GluAsnLysGluLysSerGlnIleGluIleThrArg 2297
 QY 241 CATATG-----CAGACCATCGAAGACTGTTAAAGAAAGATGACTTG 285
 DB 2298 HisLeuThrHisSerSerGlnAsnGluLysLysGluLysGluLysSerLeu 2317
 QY 286 ATGTCTGCACTAGTTTCCGTAAGAGGAGCTTG----- 318
 DB 2318 LysAspGlnIleThrAspLysSerAsnSerIleuLysCysLysGluGlnLysGlnAsn 2337
 QY 319 -----CSAGATGCGAGCAAGAAAGAGAGAGTGCCTAT 351
 DB 2338 LeuGluGluIleIleArgGlnGlnIleAlaAspIleGlnAsnSerLysPheSer---Tyr 2356
 QY 352 GAACAGCGGAACAAAGTTTGCAAATATCT-----GAGGA 387
 DB 2357 GlnIleuGluIleThrAspLysIleAsnIleSerArgLysIleuThrSerArgLysGlnIle 2376
 QY 388 GCSAAATTTTGAAGAAACCAAGCGCTTAA----- 414
 DB 2377 IleAsnMetLysGluGlnLysIleIleSerLeuLysSerGluLysGluAlaIleGln 2396
 QY 415 ATCCAGTGTGACAGTTGAGAGAGAGCTGAGAGCGAGCGAGCACTTGAAGAAAGAA 474
 DB 2397 ValAlaIleAlaGluLysArgGlnGlnHisAspLysGluIleLysGluLysGlnAsnIleu 2416
 QY 475 CTTCATCTCTCAGCAAGAGAAAGGCGCATTGAGAAAGACATGATGAAAGAGAAATTAACG 534
 DB 2417 LeuSerGlnGluGluGluGlnLysAsnIleValLeuGluGluGlnLysGlnIleValAsp 2436
 QY 535 AAAGAAAGGAGTACATGCGGATCAAGATGTGTCTGTCTGCAAGATATATGCCACTG 594
 DB 2437 LysThrAsnGlnLeuMetGlnThrLeuLysThrIleLysLysGluAsnIleGlnIle 2455
 QY 595 GAGCGCCAGGAGGAAAGGTTACAAAGAAAGATTCACAGCTATTAAT----- 642
 DB 2456 LysAlaGlnLeuAspSerPheValLysSerMetSerLeuGlnAsnAspArgAspArg 2475
 QY 643 -----CAACTGGAGAA----- 654
 DB 2476 IleValGlyAspTyrGlnGlnIleGluIleArgHisLeuSerIleIleLeuGluLysAsp 2495
 QY 655 -----ATTCAAACCCAGCTGCGCTTCGCGGAATGATGTCACAAAGGCTGTGGAGAA 708
 DB 2496 GlnLeuIleGlnIleAlaIleAlaGlnAsnAsnLysLysGluGluIleArgGly--- 2514
 QY 709 ATGCGCTATGCACTGAATAAACCAACATGAGAGAGATGAGCGAGAAAGAGCAACAGA 768
 DB 709 ATGCGCTATGCACTGAATAAACCAACATGAGAGAGATGAGCGAGAAAGAGCAACAGA 768

```

Db 2515 LeuArgSerHisMetAspSerLeuAsnHisLeuAspAlaGluLeuIle 2534
QY 769 GAGTTCAGACGAAAAAATACAGGGATCTTGAATTAAAGAT---CAGGAAATAGAGAA 825
Db 2535 GlnThrArgGlnAspLeuAsnGlnValIleThrIleLysAspSerGlnGlnLysGlnLeu 2554
QY 826 TTGAGATAGAACTGGATGAACAAACA-----CAGTTGGAACAG 867
Db 2555 LeuGlnValaGlnLeuGlnHisLysGlnLeuGlnAsnLysTyrAlaLysLeuGln 2574
QY 868 GAGCAGCAAAAGGACCCCTGGCCAGAGAGAGAGTGCCTGAGACTACAGAACTGCTGGCC 927
Db 2575 LysLeuLysGlnSerGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 2594
QY 928 GAATCTGACACCAACTG-----CAGCTC 951
Db 2595 GlnGlnLysGlnAspLeuSerLysGlnIleGlnSerLeuLysValSerIleSerGlnLeu 2614
QY 952 ACCAGATCTGAATAATAGCTCACTCACTCAAGAA----- 984
Db 2615 ThrArg---GlnValThrAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2633
QY 985 -----AAAAGTATACATATGATAATGGGAAAGTTA-----CAGAGA 1023
Db 2634 LeuLysValLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2653
QY 1024 AGAAATGAAGAAATGGAGAGACAG---TGTGTCCACAT-----CAGAG 1059
Db 2654 ArgIleAlaGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2673
QY 1060 GGGAGAGTATGACAGCATGAAGCAAGGCTAAGCCAGCTGAT----- 1104
Db 2674 GlnGlnIleGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2693
QY 1105 ---AAGCAGAGCCAGCCAGCCAGCCAGCTGCTG-----CAGCTCCTC 1146
Db 2694 MetAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2713
QY 1147 AGCAAGCAGACAGCTTCTCTG-----GAGAGCAGAGCCCTGTGCGAAGAGTG 1197
Db 2714 GlnMetGlnGlnLysLeuMetValThrLysGlnAsnLysGlnLysGlnGlnGlnGln 2733
QY 1198 GACCGGCTCGGACCCAGTTACCCAGCATGCCACATCT 1236
Db 2734 GlnSerPheGlnArgSerMetSerLeuGlnAsnSer 2746

RESULT 13
O9WVOO PRELIMINARY: PRT: 1022 AA.
AC O9WVOO:
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Sperm tail associated protein.
CN STAP OR STAP.
OS Mus musculus (Mouse).
OC Eumetazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Ohuchi J., Arai T., Kon Y., Watanabe T.;
RT "Sperm tail associated protein."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB029919; BAA82514.1; -.
KW MGD: MGI:1930136; Stap.
SQ SEQUENCE 1022 AA; 119400 MW; 4C41FE64AFCD984A CRC64;

```

```

Percent Similarity: 42.58% Conservative: 104
Best Local Similarity: 22.27% Mismatches: 158
Query Match: 9.22% Indels: 136
Db: 11 Gaps: 23

US-09-502-945-1 (1-1552) x O9WVOO (1-1022)
QY 10 GCATCCGAGAGCACTAAATCTTATGAGAAAGCTGTGAATTCAGAAATCCCAATTC 69
Db 174 AlaGlnGlnGlnLeuAlaLeuAlaGlnLysPylsIleLeuSerLeuArgSer---Leu 192
QY 70 AAGTTTTCAGACGACCTTACCTGATATTCAGAGAACTTGTGAAGTCTTAAAGGCAA 129
Db 193 AsnLeuThrArgAsp-----LysTyrGlnThrSerLeuSerAsnIle---GluLeu 208
QY 130 CTAAAGCATAAAGAAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 189
Db 209 LeuGlnCysGlnValLysMetLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 228
QY 190 TTGAATGTGCTCAGCATGAGCACTGCTTCTTCCCAACCCATACATATGTT----- 240
Db 229 GlnAsnLysGlnAspHisProLysValArgIleTyrThrSerProCysValIleGlnGln 248
QY 241 CATATCCAGACCATGCAAGACTG-----GTTAAAGAAAGATGAC 282
Db 249 HisGlnGlnThrLeuLysArgLeuSerGlnValTTPGlnLysValSerGlnGlnAsp 268
QY 283 TTGATG----- 288
Db 269 LeuIleGlnGlnLeuArgGlnLysLeuAlaCysSerAsnSerLeuValLeuGlnArgGln 288
QY 289 TCTGCATGATTTCCCTAAGAGACAGCTTGGCAGAT-----ACGACGAAAGAGAA 339
Db 289 GlnAlaLeuIleLysLeuGlnAlaGlnLysLeuIleThrAlaThrHisArgHisPro 308
QY 340 GCAAGTGTATGACAGAGTGAACAAAGTTTGCAATATTCAGAGAACCAATTTGAA 399
Db 309 ProThrSerSerGlnLysPylsGlnAspIleThrLysIleLeuLysHisLeuGlnGlnGln 328
QY 400 AAAACCAAGCTTTAATCCAGTGTGACAG-----TTGAGAAAGAGAGCTGAGAG 450
Db 329 LysAspSerGlnCysLeuHisValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 346
QY 451 CAGGCGGACGACTGAAAAAGACTTCATCTCAGCAAGAGAAAGGCCATTTGAGAA 510
Db 347 -----ArgMetClnLeuGlnAlaValIserGlnGlnLysLys---IleMetLys 362
QY 511 GACATGATGAAGAAGAAATA----- 531
Db 363 AspMetMetLysLeuGlnLeuAspLeuHisGlnLysLeuArgGlnGlnThrSerCysValIle 382
QY 532 -----ACGAAAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGAT 582
Db 383 GlnLysLysAspLysGlnThrValPheLeuGlnThrArgLeuGlnAspLeuGlnGlnGln 402
QY 583 ATTGCCCAA-----CTGAGGCCCGAGGTGGAAGGTTACAAAGCAAAAGATTTC 633
Db 403 TyrThrGlnSerGlnLysLeuSerLeuLysLysAspLysLeuGlnGlnAspLysGln 422
QY 634 GCTATTATCACTGAGAG-----GAAATTCAAAGCAGCTGCTTCTGCGGAA 681
Db 423 ArgLeuHisGlnLeuGlnLysAsnLeuMetGlnValGlnAsnSerLeuArgGlnLysGln 442
QY 682 ATGATGTTCACAAAGGTGTGTGAGAAATGCGCTATACCTGATGAATAAACCAACATGAG 741
Db 443 AlaGlnLeuGlnLys-----LeuGlnCysThrThrLysGln 454
QY 742 AAGATGAGCGCAAGAAAGAGCAGAGAGTTTCAGAGCAAAACTAACAGGATCTTGA 801
Db 455 LeuAspThrSerLeuGlnGlnLysValArgGlnSerThrSerLysIle-----AspCysGln 472
QY 802 ATTAAAGATCAGGAATATGAGAAATTTGAGAAATGAACTGATGAAAGCAACAGCTTG 861

```

Alignment Scores: 2,38e-08 Length: 1022
Score: 252.00 Matches: 114

```

Db 473 AAlaenurAgaLagLllegLlnLysLeuLysAspSerLeuGlulAlaArgSluLnu 492
Oy 862 GAGACGAGAGCAGACAGAGCCCTGGCCAGAGAGAG----- 900
Db 493 LysValSerAspGlnAsnLeuThrGlnSerLysGlulAlaHisLeuSerAlaSer 512
Oy 901 -----TGCCCTG-----AGACTAACAGAA 918
Db 513 LeuGlnAspAlaHisArgLysLllegLlnAsnLysLeuGlnAspLysGlnLysGln 532
Oy 919 CTGCTGGCGCAATCTGACACCACTGACACCACTGACACCACTGACACCACTGAC 978
Db 533 ValIleLysAspLeuGlnSerGlnLeuHisLysLeuGlnLysLysSerLysLle 552
Oy 979 CAGAAAAAGGTATACATATGATTAATTTGGAAAGTTA----- 1017
Db 553 GluGlnAlaGlyLysHisAsnArgGlnArgLeuGlnLysLeuSerSerGlnLeuSer 572
Oy 1018 CAGAGAGAA-----AATGAAGAAATTGGAGAGAACAGATGTCTCCAGACATG 1071
Db 573 GlnArgArgLeuSerAsnAlaGlnLysGlnLysSerLeuLeuGln----- 587
Oy 1072 GAGACGATGAAGCAAGGCTTAAGCAGCTGATTAAG-----CACAGCCAGGCC 1125
Db 588 LysThrLeuAspGlnAspGlnLysLysLysLysLysLysLysLysLysLys 607
Oy 1126 CAGCAG-----CTGGTGACGCTCCACACAGCAGAACACACCTTCTCTGAG 1173
Db 608 GluGlnLysGlnArgGlnLeuThrAsnSerLleArgLysLeuGlnGlnLysLeu 627
Oy 1174 -----AGCGAG 1179
Db 628 IleLysGlnLeuLeuGlnGlnLysArgGlnGlnLysLysSerLysGlnGlnLys 647
Oy 1180 AGCTGTGCGAAGAGGTGACCGCTGCGAGCCAG 1215
Db 648 AlaLeuGlnGlnLysLllegLlnAlaLeuArgGlnGln 659

RESULT 14
Oy 99BM8 PRELIMINARY; PRT; 876 AA.
ID 09BM8
AC 09BM8;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE Paramyosin.
OS Sarcophaga scabiei.
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
OC Acariformes; Sarcopitiformes; Astigmata; Sarcopitidae; Sarcopitidae;
OC Sarcophages.
OX NCBI_TaxID=52283;
RN [1]
RP SEQUENCE FROM N.A.
RA Mattsson J.G., Ljunggren E.L., Bergstrom K.;
RT "Paramyosin from the parasitic mite Sarcophaga scabiei: cDNA cloning
RT and heterologous expression.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF317670; AA01181.1;
DR InterPro; IPR002928; Myosin_tail.
DR Pfam; PF01576; Myosin_tail; 1.
SQ SEQUENCE 876 AA; 102454 MW; C99475EE7A0DAE52 CRC64;

Alignment Scores:
Pred. No.: 2.7e-08
Score: 251.00
Percent Similarity: 43.33%
Best Local Similarity: 23.85%
Query Match: 9.19%
DB: 5 Gaps: 17

US-09-502-945-1 (1-1552) x 09BM8 (1-876)
Oy 16 GAGAGCTAAACTTACTTATGAGAGAAAGTGTCAA-----ATTGAGCAA 60

```

```

Db 374 GluLysLeuAsnLeuAspLeuLysAsnLysLeuGlnGlnValThrMetLeuMetGln 393
Oy 61 TCCCAATGAAAGTTTGTGAGAACACCTTACGTGAATTCAGAGAACTTGAGAAATCTT 120
Db 394 AlaGln-----LysGlnAlaArgAlaLysAlaAlaGlnLeuGlnHisLysLys 412
Oy 121 AAAGCAACATTAAGCATTAAGAAATTTCTGCTGCTGCTGAATTAATCTGTAACCGTGTGT 180
Db 413 ----GluLysLeuArgAspGlnArgAspAlaLeuAlaArgGlnAsnLysLeuThrAsp 431
Oy 181 GCTCTTTGTTGAATTTGCTCCAGCATGAGCTGCTTTTCCCAACCATTAATGTT 240
Db 432 AspLeu-----AlaGlnLysSerGlnLeuAsnAspAlaHisArgArgLle 447
Oy 241 CATATGCAGACCATTCGAA-----AGACTGGTTAAAGAAAGATGATGATGCTGCA 294
Db 448 HisGlnGlnLysLllegLlnLysArgLeuGlnAsnGlnArgGlnLysLeuSerAla 467
Oy 295 CTAGTTCCGTAAGAGCAGCTTGCCAGATACG-----CAGCAAGAGAGCAGAGTCT 348
Db 468 TyrLysGln-----AlaGlnThrLeuArgLysGlnGlnLysAlaAsn 482
Oy 349 TATGAACAGGTGAAACAAGTTTGCATAATCTGAGAGAACCAATTTGAAAAACCAAG 408
Db 483 GlnArgLeuThrAlaGlnLysLeuAlaArgHisAspTyrGlnLysArgLeuAlaGln 502
Oy 409 GCTTAATTCAGCTGACACAGCTTGAGAGAGAGCTGAGAGAGCAGCGAGCTTGAA 468
Db 503 LysGlnGlnGlnLllegLlnAlaLeuArgLysGlnTyrGlnLllegLlnLysLeuAsn 522
Oy 469 AAAGAACTTGCACTCAGCAAGAGAAAGGCGCATTTGAGAAAGCATGAGTGAAGAA--- 525
Db 523 MetArgLeuAlaGlnAlaGlnLysLeuLysThrGlnLleAlaArgLeuLysLys 542
Oy 526 -----GAAATPACGAAA-----GAAAGGAGTACATG 552
Db 543 TyrGlnAlaGlnLllegLlnLysLeuLysSerLeuAspAlaAlaAsnLysAlaAsn 562
Oy 553 GGATCAAGATGTTGATCTGCTCAG-----AATATTGCCCACTGAGGCCAGCGTG 606
Db 563 AspLeuGlnLysThrLleLysLysGlnAlaLeuGlnLllegLlnLysLysLys 582
Oy 607 GAAAGGTTTACAAAGGAAAGATTTCAAGTATTAATCAAGGAGAAATTCAAAGCCAG 666
Db 583 AspGlnValHisArgGlnLeuGlnGlnAlaValAspGlnLeuGlyValThrGlnArg 602
Oy 667 CTGCGCTTCGGAAGATGATGTCACAAAGGTGTGTGAGAAATGCGTATCAGCTGAT 726
Db 603 CysGlnAlaLeuGlnAlaGlnLysLeu-----GlnGlnArgGlnAlaLeuGln 618
Oy 727 AAACCAACATGAGAGAGATGAGCGAGAAAGAGCAGCAGAGATTCAGAGCAAAACT 786
Db 619 GlnAlaAsnArgAlaLysArgGlnAlaGlnLysLeuHisGlnGlnAlaValAlaArg 638
Oy 787 AAC----- 789
Db 639 AsnGlnLeuThrThrIleAsnValAsnLeuAlaSerAlaLysSerLysLeuGlnSer 658
Oy 790 -----AGGATCTTGAATTAAGATCAG 813
Db 659 PheAlaAlaLeuGlnAsnAspTyrAspLysGlnLysGlnLysGlnLysLysLys 678
Oy 814 GAAATAGAGAAATTGAGAAATGAGACTGATGAGAAACAAACACTTGGAACAGAGCAG 873
Db 679 ArgValGlnLysLeuThrLllegLlnLysSerThrLysAspLeuValGlnGlnGln 698
Oy 874 CAGAGGAGAGCGCTGCGCAGAGAGAGGAGTCCAGACATCAACAGAACTCTGGGCAACT 933
Db 699 GlnArg-----LeuValLysMetGlnThrValLys-----LysSerLeuGlnLys 714
Oy 934 GAGCACAACGACCTCAGCAGATCTGAATAGCT-----CAACTGATCAAGAAAAA 987

```



```

QY 1051 ---GTCAGCATGGAGAGTACATGAGACGATGACCAAGCCTAAGCAGCTGATAG 1107
      ::::: ||| ::||| ||::: ||| :::
Db 639 GluLeuGluGluLeuArgAlaMetValGluGlnSerGluArgAlaArgLysLeuAlaGlu 658
QY 1108 CACAGCCAGGCCACAGCCACAGCTGTGTCAGCTCTCTCAGCAGAGAAC---CAGCTT 1164
      ||| ::| ||||| ||| ||||| |||
Db 659 GlnGluLeuIleGluAlaSerGluArgValGlnLeuLeuHisSerGlnAsnThrSerLeu 678
QY 1165 CTCCTGGAGAGCGAGACCTGTGCGAGAGAGGTGACCGGCTGCGGAGACCACTTACCAGC 1224
      ::| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Db 679 IleAsnGlnLysLysMetGluAlaAspIleSerGlnLeuGlnThrGluValGluGlu 698
QY 1225 ATGCACACATCTGATTGC 1242
      ||| ::::: |||
Db 699 AlaIleGln---GluCys 703

```

Search completed: March 21, 2003, 12:57:58
 Job time : 81.2241 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 11:21:37 ; Search time 14.6271 Seconds

(without alignments)

8801.668 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732

Sequence: 1 cctctgcatcgcacgcagaaa.....aaatgaacttttaagaaga 1552

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame_n2p.model -DEV-xlp
-Q/cg2_1/USPTO.spool/US09502945/rnauc_14032003_101058_19100/app_query.fasta_1.10979
-DB-SwissProt_40 -QFMT-fastan -SUFFIX-rsp -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -DIST=45
-DOCALLGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE-LOCAL
-OUTFMT=pro -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09502945_6CGN_1_1_113_6runat_14032003_101058_19100 -NCP=6 -ICPU=3
-NO_XLPEX -NO_MMAP -LARGEOUTPUT -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEDOUT=120
-WARN_TIMEDOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	264	9.7	3911	AKA9_HUMAN	Q99996 h a-kinase
2	256.5	9.4	1790	USO1_YEAST	P25386 saccharomyc
3	255.5	9.4	978	RASO_AQUAE	O67124 aquifex aeo
4	255.5	9.4	1959	MYH9_CHICK	P14105 gallus galli
5	254	9.3	2017	MYH9_CHICK	Q99323 dirosophila
6	248.5	9.1	4687	PLEL_RAT	P30427 rattus norv
7	245	9.0	2245	MYH3_DICDI	P54697 dictyostel
8	244	8.9	1940	MYH3_CHICK	P02565 gallus gall
9	244	8.9	4684	PLEL_HUMAN	O15149 homo sapien
10	243	8.9	1960	MYH9_HUMAN	P35579 homo sapien
11	242	8.9	2663	CENE_HUMAN	O02224 homo sapien
12	241.5	8.8	1755	PEPL_MOUSE	O91269 mus muscul
13	241.5	8.8	1938	MYH8_HUMAN	P35580 homo sapien
14	240.5	8.8	1976	MYH8_HUMAN	O01202 brugia mala
15	239.5	8.8	880	MYSP_BRUMA	O02171 onchocerca
16	239	8.7	879	MYSP_ONCVO	P13332 dirofilaria
17	238	8.7	848	MYSP_DIRIM	P13335 homo sapien
18	237	8.7	1937	MYH8_HUMAN	

19	236.5	8.7	1934	1	MYH7_MESAU	P13540 mesocricetu
20	236	8.6	1935	1	MYH7_RAT	P02564 rattus norv
21	236	8.6	1940	1	MYH8_RAT	P12847 rattus norv
22	235.5	8.6	1976	1	MYH8_BOVIN	Q27991 bos taurus
23	235	8.6	1130	1	YL17_CAEL	Q11102 caenorhabdi
24	235	8.6	1972	1	MYH8_RABIT	P35748 oryctolagus
25	234.5	8.6	1939	1	MYH4_HUMAN	O95623 homo sapien
26	234	8.6	1549	1	TRHY_SHEEP	P22793 ovis aries
27	234	8.6	1756	1	PEPL_HUMAN	O60437 homo sapien
28	234	8.6	1935	1	MYH7_PIG	P79233 sus scrofa
29	234	8.6	1961	1	MYH9_RAT	O62812 rattus norv
30	234	8.6	4473	1	PLEL_CRGR	O91155 cricetus
31	233.5	8.5	794	1	HMMR_MOUSE	O00547 mus musculu
32	233.5	8.5	1433	1	RESE_CHICK	O42184 gallus galli
33	233.5	8.5	1818	1	HMM2_MYCPN	P75471 mycoplasma
34	233.5	8.5	1935	1	MYH7_HUMAN	P12863 homo sapien
35	233.5	8.5	1935	1	MYSS_CYPCA	O90339 cyprinus ca
36	233.5	8.5	1938	1	MYH_AEOIR	P24733 aequipecten
37	232.5	8.5	976	1	SCPI_HUMAN	O15431 homo sapien
38	232.5	8.5	1087	1	AKA9_RABIT	O28628 oryctolagus
39	232.5	8.5	1939	1	MYH1_HUMAN	P12882 homo sapien
40	232.5	8.5	1957	1	YD86_SCHPO	O10411 schizosacch
41	232.5	8.5	1976	1	MYH8_RAT	O91170 rattus norv
42	232	8.5	1938	1	MYHD_HUMAN	O9ukx3 homo sapien
43	232	8.5	1972	1	MYH8_MOUSE	O08638 mus musculu
44	231.5	8.5	2704	1	BPA1_HUMAN	O03001 homo sapien
45	231.5	8.5	3210	1	CENF_HUMAN	P49454 homo sapien

ALIGNMENTS

RESULT 1

AKA9_HUMAN STANDARD; PRT; 3911 AA.
ID Q99996; Q9U004; Q9U003; Q9Y6T2; O14869; O43355; O94895; Q9Y6B8;
AC G99996; Q9U004; Q9U003; Q9Y6T2; O14869; O43355; O94895; Q9Y6B8;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
DE (PRAK9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
DE (hyperion protein) (Yotiao protein) (Centrosome- and golgi-Localized
PRN-associated protein) (CG-NAP).
GN AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP TISSUE=Brain;
RC MEDLINE=98151389; PubMed=9482789;
RA Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.,
Yotiao, a novel protein of neuromuscular junction and brain that
interacts with specific splice variants of NMDA receptor subunit
NR1.
RT J. Neurosci. 18:2017-2027(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RX MEDLINE=99219864; PubMed=10202149;
RA Witczak O., Skalniak B.S., Keryer G., Bornens M., Tasken K.,
Jahansen T., Oestavik S.;
RT "Cloning and characterization of a cDNA encoding an A-kinase anchoring
protein located in the centrosome, AKAP450.";
RL EMBO J. 18:1858-1868(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Brain;
RX MEDLINE=99287934; PubMed=10358086;
RA Takahashi M., Shibata H., Shimokawa M., Miyamoto M., Mukai H., Ono Y.,
RT "Characterization of a novel giant scaffolding protein, CG-NAP, that
anchors multiple signaling enzymes to centrosome and the golgi
apparatus.";

FT CONFLICT 3833 3833 T -> S (IN REF. 3).
 SQ SEQUENCE 3911 AA; 453664 MW; 3FB1CB1C819B47AA CRC64;
 Alignment Scores:
 Pred. No.: 1.32e-07 Length: 3911
 Score: 264.00 Matches: 121
 Percent Similarity: 40.57% Conservative: 107
 Best Local Similarity: 21.53% Mismatches: 162
 Query Match: 9.66% Indels: 172
 DB: 1 Gaps: 22
 US-09-502-945-1 (1-1552) x AK9_HUMAN (1-3911)
 QY 1 CTTGTGATGATCCGAGAGCTTAACCTTACTTATGAGAAAGGTGAA---ATTGAG 57
 Db 1855 MetLeuAnIleSerSerThrLeuGlnAlaValGluLysLeuLeuGlnAlaIleSer 1874
 QY 58 GAA-----TCCCAATTGAACTTTTGGAGAACGAC 87
 Db 1875 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetIrrgIuSer 1894
 QY 88 TTAGCTGATATCAGAGAACTTGAAGCTTTAAAGCAACTAAAGCATTAAGAATT 147
 Db 1895 PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysCysGlnGlnGlnLeuArgIuArg 1914
 QY 148 CTTGTGCTGCTAATCTGTAAACCGTGTGGTGGCTTTGTTGAATGTGTCAGCAT 207
 Db 1915 LeuHisGlnGlnSerArgAlaArgIu-----GlnLeuAlaValGlnLeuSerLysAla 1932
 QY 208 GAAGCTGCTTTCCCAAAACCCATCTAATGTCATATGACAGCCATGAAAGACTGTT 267
 Db 1933 GluGlnValIleAspGlyTyrAlaAspGlnLysThrIleu-----PheGlnArgGlnIle 1950
 QY 268 AAAGAAAGATGATGATGATGCTCTGCACTA-----GTTTCCGTAAAG 309
 Db 1951 GlnGlnLysThrAspIleLeuAspArgLeuGlnGlnLeuLysAlaSerAsnArg 1970
 QY 310 AGCACTTGGCAGATCCGACCAAGA----- 336
 Db 1971 LeuGlnGlnLeuGlnAlaGlnGlnGlnGlnIleGlnGlnGlnArgGlnLeuLeuSerArg 1990
 QY 337 -----GAGCAAGTGCTTATGAACAGCGTGAACCAAGCTTTTGCAG 375
 Db 1991 GlnLysGlnAlaMetLysAlaGlnAlaGlnIleProValGlnGlnIle-----GlnLeuLeuGln 2008
 QY 376 ATATCTGAGGAAGCAATTTGAAGAAACAGGCTTATATCCATGCTGACCACTTGAG 435
 Db 2009 GluThrGlnLysLeuMetLysGlnLysLeuGlnValGlnCysGlnAlaGlnValArg 2028
 QY 436 AAGGAGCTGAGAGCGAGCGGAGCTTGAAGAAAGAACTGCAATCAG----- 486
 Db 2039 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnIleValSerArg 2048
 QY 487 -----CAAGAGAAAGGGCCATTGAGAAAGACATG----- 516
 Db 2049 PheIleGlnLeuGlnGlnGlnLysAsnThrGlnLeuMetAspLeuArgGlnGlnAsnGln 2068
 QY 517 ---ATGAAAGAGAAATACCAAGAAAGGAGTACATGGATCAAGATGTTG----- 567
 Db 2069 AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2088
 QY 568 -----ATCTGTCTCAGAAATATGGCCCAACTGGAGCCCAAGTGGAGAAAG 612
 Db 2089 GlnHisGlnArgAspValPheGlnGlnGlnIleGlnLysGlnGlnGlnLeuLysVal 2108
 QY 613 GTTACAAAGAAATATTACATTAATCAACTGAGAGAAATTAAGCAAGCTGAGT 672
 Db 2109 ValProAlaArgPheGlnProIleSerGlnHisGlnThrArgGlnValGln-----GlnLeuAla 2127
 QY 673 TCTCGGAAATGATGCACAAAGGTGTGTGAGAAATG-----CGTATACAGCTG 723
 Db 2128 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnIleu 2147

QY 724 AATAAACCAATGAGAGAGATGAGCGAGAAAGAGCACAGAGATTCAGACCAAA 783
 Db 2148 GlnIrrgAspIleGlnGlnIrrgAsnGlnu---GlnIleGlnLysLeuGlnIrrgValArg 2166
 QY 783 ----- 783
 Db 2167 GlnLeuGlnGlnAlaLeuLeuValSerAlaAspThrPheGlnLysValGlnLysAspArgLys 2186
 QY 783 ----- 783
 Db 2187 HisPheGlnAlaValGlnAlaLysProGlnLeuSerLeuGlnValGlnLeuGlnAlaGln 2206
 QY 784 -----ACTAAC----- 789
 Db 2207 ArgAspAlaIleAspArgLysGlnLysGlnLysGlnIleThrAsnLeuGlnGlnIleuGln 2226
 QY 790 -----AGGATCTTGAATTAAGATTCAGAAATAGAAATTAAGATTAAGATGAT 843
 Db 2227 PheArgGlnGlnLysGlnLysAsnGlnGlnValGlnGlnIleuHisMetGlnLeuGln 2246
 QY 844 -----GAAAGCAACACACTTGAACAGAGACGACGACAGAGCCCTGGCC 891
 Db 2247 IleGlnLysLysGlnSerThrThrArgLeuGlnGlnGlnGlnLysLysPhe 2266
 QY 892 AGAGAGAGTGGCTGAGACTACAGAACTGCTGGCGCAATGAG-----CAC 939
 Db 2267 LysAspAspMetGlnLysLeuGlnLysLeuAlaIleLysGlnSerAspAlaMetSerThrGln 2286
 QY 940 CAACGTGACCTGCACAGATCTGAATAGCTCAACTCACTCAAGAAAGAAATATACATAT 999
 Db 2287 AspGlnHisValLeuPheGlnLysPheAlaGlnIleIleGlnGlnLysGlnValGlnIle 2306
 QY 1000 GATAAATGGGA-----AAGTTACGAGAAAGA-----AAT 1029
 Db 2307 AspGlnLeuAsnGlnGlnValThrLysLeuGlnGlnIleLysIleThrThrAspAsn 2326
 QY 1030 GAAGAAATGGAGAA----- 1044
 Db 2327 LysValIleGlnGlnLysAsnGlnLeuIleArgAspLeuGlnThrGlnIleGlnLysLeu 2346
 QY 1045 -----CACTGTGTCAGCATGGAGAGTACATGACAGCATGACCAAGGCTA 1092
 Db 2347 MetSerAspGlnGlnLysValLysArgAsnArg-----GlnGlnIleuIle 2361
 QY 1093 AGCAGCTGAGTAACACAGACCCAGCCACAGCCAGCTGTCACCTCTCAGCAAG 1152
 Db 2362 GlnGlnLeuAsnGlnValIleGlnLysLeuGlnGlnGlnLeuAlaAsnIleGlnLys 2381
 QY 1153 CAGAACCAAGCTTCTCTGAGAGCGACAGACCTGTGCGAAGAGTGGACCGTGGCAGC 1212
 Db 2382 ThrSer-----MetAsnAlaHisSerLeuSerGlnGlnAlaAspSerLeuLysHis 2398
 QY 1213 CAGTTA 1218
 Db 2399 GlnLeu 2400
 RESULT 2
 USOL_YEAST STANDARD: PRT, 1790 AA.
 AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 GN USOL OR INT1 OR YDL058W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=X2180-1A;
 RX MEDLINE=91185402; PubMed=2010462;

RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usol, is required for intracellular
RT protein transport in *Saccharomyces cerevisiae*.";
RL J. Cell Biol. 113:245-260(1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McCellan M., Tao N.,
RA Kendrick K.E.;
RL Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANS. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOL/YBL047C FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X54378; CAA38253.1; -;
DR EMBL: L03188; AAB00143.1; -;
DR EMBL: U53668; AAB66659.1; -;
DR PIR: A38455; A38455.
DR SGD: S0002216; USOL.
DR Interpro: IPR002017; Spectrin.
KW Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724 GLOBULAR HEAD.
FT DOMAIN 725 1790 COILED COIL (POTENTIAL).
FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
FT DOMAIN 1172 1786 ASF/GLU-RICH (ACIDIC).
FT CONFILCT 847 847 G -> E (IN REF. 2).
FT CONFILCT 924 924 E -> K (IN REF. 2).
FT CONFILCT 1253 1253 V -> I (IN REF. 2).
FT CONFILCT 1319 1319 I -> V (IN REF. 2).
FT CONFILCT 1461 1461 N -> S (IN REF. 2).
FT CONFILCT 1581 1581 G -> S (IN REF. 2).
FT CONFILCT 1600 1600 I -> V (IN REF. 2).
FT CONFILCT 1661 1661 R -> S (IN REF. 2).
FT CONFILCT 1772 1772 D -> DEEDDEE (IN REF. 2).
SO SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CMC64;

Alignment Scores:

Pred. No.: 3.72e-07 Length: 1790
Score: 256.50 Matches: 86
Percent Similarity: 43.33% Conservative: 99
Best Local Similarity: 20.14% Mismatches: 169
Query Match: 9.39% Indels: 73
DB: 1 Gaps: 11

US-09-502-945-1 (1-1552) x USOL_YEAST (1-1790)

OY 64 CAATGAGCTTTTGAAGCAGCTTAGCATATCAGAGAACTTGTGAAGATCTT--- 120
Db 963 LysLeuLysSerLeuAlaAsnAsnTrpLysAspMetGlnAlaGluAsnGluSerLeuIle 962
OY 121 -----AAGAGCACTAAAGCACTAAAGATTTCTTGCTGCTCTAATATCTTAACCGT 174
Db 983 LysAlaValGluGluSerLysAsnGluSerSerIleGlnLeuSerAsnLeuGlnAsnLys 1002

OY 175 GTTGTGGTCTTTTGTGAATGTCTCAGCATGAGAGCTGTCTTCCCAACCCACTACT 234
Db 1003 IleAspSer-----MetSerGlnGluGlu 1011
OY 235 AATGTTATATGAC-----ACATCGAAGACTGGTTAAAGAAAGA 276
Db 1012 AsnheGlnIleGluArgGlySerIleGluLysAsnIleGlnGluLysTrpIle 1031
OY 277 GATGACTGTGATGTGCACATGTTCCGTAGGACAGCTGGCAGATPACGACCAAGA 336
Db 1032 SerAspLeuGluGlnThrLysGluGluIleLysSerLysSerAspSerLysAspGlu 1051
OY 337 GAACCAAGTGTATGACAGGTGAACCAAGTTTGCATATCTGAGAGCAACCAATTT 396
Db 1052 TyrGluSerGlnIleSerLeuLeuGluGluLysLeuGluThrAlaThrAlaAsnAsp 1071
OY 397 GAAAAACCAAGGCTTTAATCCAGTGTGACCACTTGAGAGAGAGCTGGAGAGCAG--- 453
Db 1072 GluAsnValAsnLysIleSerGluLeuThrLysThrArgGluGluGluAlaGluLeu 1091
OY 454 -----GGCAGCGACTT 465
Db 1092 AlaAlaTrpLysAsnLeuLysAsnGluLeuGluThrLysLeuGluThrSerGluLysAla 1111
OY 466 GAAAAAGACTTGATGATCTCAGCAGAGAGAGAGGCACTTGAGAGAGAGAGAGAG 525
Db 1112 LeuLysGluValLysGluAsnGluGluGlnIleLysGluGluLysIleGlnGluLys 1131
OY 526 GAATATACGAAGAAGAGAGAGCTACATGAGATCAAGATGTTGATCTGTCAAAATAT 585
Db 1132 GluAlaThrGluThrLysGluGlnLeuAsnSer----- 1142
OY 586 GCCCACTGAGAGGCCAGTGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 645
Db 1143 -----LeuArgAlaAsnLeuGluSerLeuGluGluGlnIleLysGluAsnLeuAlaGln 1160
OY 646 CTGAGAGAGAAATTCAGACCCAGCTGGCTTCTCGGAGAAATGCATGCACAAAGCTGTGGA 705
Db 1161 LeuLysLysTrpArgGluGlnGluIleAlaAsnLysGluArgGlnTrpAsnGluGluIleSer 1180
OY 706 GAATATGGCTATGCGATGATTAACCAACATGAG-----AAG 744
Db 1181 GlnLeuAsnAspGluIleThrSerThrGlnGlnGlnGluAsnGluSerLysLysAsn 1200
OY 745 GATGAGCAGAT 804
Db 1201 AspGluLeuGluGluGluValLysAlaMetLysSerThrSerGluGluGlnAsnLeu 1220
OY 805 AAGATCAGAGAAATAGAGAAATTTGAGATAGACTGATGAGAGAGAGAGAGAGAGAG 864
Db 1221 LysLysSerGlnIleAspAlaLeuAsnLeuGlnIleLysGluLeuLysLys-----Lys 1238
OY 865 CAGAGCAG 906
Db 1239 AsnGluThrAsnGlnAlaSerLeuLeuGluSerLysSerValGluSerGluThrVal 1258
OY 907 AGACTTAACAGAT 966
Db 1259 LysIleLysGluLeu-----GlnAspGluLysAsnPhenylGluLysGluVal 1274
OY 967 GCTCACTCAGTCAAG 1017
Db 1275 SerGluLeuGluLysAsnLysLysLysLysLysLysLysLysLysLysLysLys 1294
OY 1018 CAG 1077
Db 1295 GlnLysGluSerGluLysIleLysGluLeuAspAlaLysThrArgLeuLysIle 1314
OY 1078 ATGAGCAAG 1137
Db 1315 GlnLeuGluLysIleThrAsnLeuSerLysAlaLysGluLysSerGluSerGluLeuSer 1334
OY 1138 CAGCTCCTCAGCAG 1197

[illegible][illegible]


```

Db 1207 Gln-----ThrLysArgValLysAlaAsnLeuGluLys-----AlaLysGlnIle 1221
QY 529 ATACGAAAGAAAGGAGTACATGGATCAAAAGATGTTGATCTTCTCAGAAATATTCC 588
Db 1222 LeuGlnSerGluArgAlaGlnLeuSerAsnGlnValLysValLeuGlnGlnLysGly 1241
QY 589 CACTGAGGCGCCAGGTGGAAAGATTACAAAGAAAGAAATTCGATTAATCAACTG 648
Db 1242 AspaLagLuuHsLys-----ArgLysLysValAspAla-----GlnLeu 1254
QY 649 GAGGAATATCAAGACCGACCTGCTCTCGGGAAGATGTCACAAAGCTGTGTGAGAA 708
Db 1255 GlnGluLeuGlnValLysPheThrGluGluLys-----ArgValLysThrGlu 1270
QY 709 ATGCCCTATCAGCTGATTAACCAACATGGAG----- 741
Db 1271 LeuAlaGluArgValAlaSnLysLeuGlnValGlnLeuAspAsnValThrGlyLeuLeuAsn 1290
QY 741 ----- 741
Db 1291 GlnSerAspSerLysSerIleLysLeuAlaLysAspPheSerAlaLeuGlnSerGlnLeu 1310
QY 742 AAGATGAGCGAGAAAGAGACACAGAGATTCAAGAGCAAAACAACTAAGCGATCTTGA 801
Db 1311 GlnAspThrGlnGlnLeuLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1330
QY 802 ATTAAGATCAGCAAAATAGACAATTTGAGATA-----GAACCTGGATGA 846
Db 1331 LeuLysGlnThrGlnLysAspGlnLysAsnAlaLeuLysGlnGlnGlnGlnGlnGln 1350
QY 847 AGCAACACACTGGAGACAGAG-----CAGCAAGAGCGCCCTGGCCGA 894
Db 1351 AlaLysArgAsnLeuGlnLysGlnLysLeuSerValLeuGlnGlnGlnAlaValGlnAlaArg 1370
QY 895 GAGGAG-----TGCCCTGAGACTAACAGCAACTGCTGGCGAATCTGAG 936
Db 1371 LysLysMetAspAspGlyLeuGlnLysLeuGlnLysLeuGlnLysLeuGlnLysLeu 1387
QY 937 CACCAACTGCACCTCACCAGATCTGAAATAGCTCACTCAAGCAAGAAAGGTATACA 966
Db 1388 LysLysLeuGlnLysAspLeuGlnSerLeuThrGlnArgTyrGlnGlnGlnLysLeuAla 1407
QY 997 TATGATTAATTTGGGAAATCTTACAGAGAGA---AATGAAGAAATGGAGAACAGCTGTC 1053
Db 1408 TyrAspLysLeuGlnLysThrLysThrArgLeuGlnGlnLysLeuAspPheLeuAlaVal 1427
QY 1054 CAGCATGGAGATGATCATGAGACGATGAAGCAAGGCTAAGCAGCTGATTAAGCAGAC 1113
Db 1428 -----AspLeuAsp---HisGln 1432
QY 1114 CAGGCCACAGCCACAGACCTGCTGCTCCTCAGACAGCAAGCAACACCTTCTCTGGAG 1173
Db 1433 ArgGlnThrValSerAsnLeuGlnLysLysGlnLysLysPheAspGlnLeuAlaGln 1452
QY 1174 AGCAGAGCTCTGTCG-----GAGCAGAGTACGCGCGGAGACCCAG 1215
Db 1453 GluLysAsnIleSerAlaLysTyrAlaGlnGlnLysArgAspArgAlaGlnAlaGln 1470

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=90349606; PubMed=21117279;
RA Kechum A.S., Stewart C.T., Stewart M., Klehart D.P.;
RT Complete sequence of the Drosophila nonmuscle myosin heavy-chain
RT transcript: conserved sequences in the myosin tail and differential
RT splicing in the 5' untranslated sequence."
RL Proc. Natl. Acad. Sci. U.S.A. 87:6316-6320(1990).
CC -1- FUNCTION: NONMUSCLE MYOSIN APPEARS TO BE RESPONSIBLE FOR
CC CELLULARIZATION. REQUIRED FOR MORPHOGENESIS AND CYTOKINESIS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM: ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M35012; AAA28713.1; -.
CC PIR: A36014; A36014.
CC PIR: B36014; B36014.
CC HSSP: P10587; IBR2.
CC FLYBase: FBgn0005634; zip.
CC InterPro: IPR000048; IQ_region.
CC InterPro: IPR004009; Myosin_N.
CC InterPro: IPR002928; Myosin_tail.
CC InterPro: IPR001609; myosin_head.
CC Pfam: PF00612; IQ; 1.
CC Pfam: PF01576; Myosin_tail; 1.
CC Pfam: PF02736; Myosin_N; 1.
CC PRINTS: PR00193; MYOSINHEAVY.
CC ProDom: PD000355; myosin_head; 1.
CC SMART: SM00015; IQ; 1.
CC SMART: SM00242; MYSC; 1.
CC PROSITE: PS50096; IQ; 1.
CC Myosin: Alternative splicing: Colled coil; Actin-binding;
CC ATP-binding; Calmodulin-binding;
CC MYOSIN HEAD-LIKE.
CC DOMAIN 1 829
CC FT 830 859
CC FT DOMAIN 886 2017
CC FT NP_BIND 225 232
CC FT DOMAIN 250 260
CC FT DOMAIN 682 694
CC FT DOMAIN 705 727
CC FT DOMAIN 742 758
CC FT DOMAIN 1303 2017
CC FT DOMAIN 1303 1970
CC FT DOMAIN 1971 2017
CC FT VARSPLIC 1 45
CC FT SEQUENCE 2017 AA; 232016 MW; 73E3CB02BA8F2528 CRC64;

```

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	4.98e-07	254.00	110	101	174	104
Percent Similarity:		43.15%				
Best Local Similarity:		22.49%				
Query Match:		9.30%				

US-09-502-945-1 (1-1552) x MYSN_DROME (1-2017)

```

QY 4 CTGATGATCCGAGAACCTTAACCTTATATAGCAAAAGTGTGAATTTGAGGAATCC 63
Db 1143 IleAspGlnGlnSerAlaThrLysAlaGlnLysAlaGlnArgGlnLysGlnLys 1162
QY 64 CAATTGAAGTTTGGAGAACGACTTA-----GCTGAATATTCAG 102

```

```

Db 1163 GlnLeuLacIuIleGInGluAspLeuGluAgluLysAlaAlaArgAlaLysAlaGlu 1182
QY 103 AGAAGTCGTGAAGATCTTAAAGACCACTAAAG-----CATAAAGATTTCTTCGGCT 156
Db 1183 LysValAlaArgArgAspLeuSerGlnuLueGluLalaLeuLysAsnGluLeuLeuAspSer 1202
QY 157 GCTAAATCTGTGAACCGCTGTGGTGTCTTGTGTAATGTGCTCAG-----204
Db 1203 LeuAspThrAlaAlaGInGInGluLeuAlaArgSerLysArgGInGInGluLeuAlaThr 1222
QY 205 -----CATGACGTGTTCTT-----219
Db 1223 LeuLysLysSerLeuGInGluGluThrValAsnHisGluLysValLeuAlaAspMetArg 1242
QY 220 -----TCCCAACCCACTACTATGTCATATGCAGACACTGAAAGCTGTGTA 270
Db 1243 HisLysHisSerGInGInGluLeuAsnSerIleAsnAspGln-----LeuGluAsnLeuArgLys 1261
QY 271 -----GAAAGAGATGACTGTGATG 288
Db 1262 AlalysThrValLeuGluLysAlaLysGlyThrLeuGluLalaGluAsnAlaAspLeuAla 1281
QY 289 TCTGCACACTATTCCTTAAAGAGACGCTTGCGACATACGACGAAAGAGCAAGTGTCT 348
Db 1282 ThrGluLeuArgSerValaAsnSerSerArgGInGluAsnAspArgArgLysGlnAla 1301
QY 349 TATGACAGAGGTGAACAAAGTTTGCACAAATCTGAGAGAACCAATTTGAAAAACCAAG 408
Db 1302 GluSerGlnIleAlaGlu-----LeuGlnVal-----LysLeuAlaGluIleGluAlaArg 1319
QY 409 GCTTTAATCCAG-----TGTGACCACTGAGAGAGCGTGCAGAGCGCGAGCGA 462
Db 1320 SerGluLeuGInGInLysGlyCysThrLysLeuGInGInGluLalaAsnIleThrAsnGln 1339
QY 463 CTGGAAGAAA-----GACTTGCACTCTGACGAAGAAA 495
Db 1340 LeuGInGluLalaGluLeuLysAlaSerAlaAlaValLysSerAlaSerAsnMetGluSer 1359
QY 496 AGGCGCATGAGAAAGACATGATGATAAGAAATACGAAAGAAAGAGGATCATGTGGA 555
Db 1360 GlnLeuThrGluLalaGInGInLeuLeuGInGluGluThrArgGlnLys-----1375
QY 556 TCAAAAGATGTTGATCTGTCTCAGAAATATGCCCACTGAGGCCCGAGTGAAGAGTT 615
Db 1376 -----LeuGlyLeuSerLysLeuArgGlnIleGluSerGluLysGluAlaLeu 1392
QY 616 ACAAGAGAAAGATTTACGATTAATACACTGAGGAAATTCAAAGCCAGTGGCTTCT 675
Db 1393 GlnGlu-----GlnLeuGInGluLysAspGlnAlaLysArgAsn 1405
QY 676 CGGGAATGATGTCTACAAAGGTGTGTGAGAAATGCGTATCAGCTAATTAACCAAC 735
Db 1406 TyrGluArgLysLeuAlaGluValThrThrGlnMetGln-----GluIleLysLysLysAla 1424
QY 736 ATGAGAGAGATGAGCGAGAAAAGAGACACAGAGAGTTCCAGACCAAACTAACAGAGAT 795
Db 1425 GlnGluAspAlaAspLeuAlaLysGluLeuGInGluLysLysArgLeuAsnLysAsp 1444
QY 796 CTGGAATTAATTAATCAGAAATAGAGAAATTCAGATTAAGAA-----CTGGATGAA 846
Db 1445 IleGluAlaLeuGluLysGlnValLysGluLeuIleAlaGlnAsnAspArgLeuAspLys 1464
QY 847 AGCAAAACACTTGGAGACGAGACGACAGACAGCC-----885
Db 1465 SerLysLysLysIleGlnSerGluLeuGluAspAlaThrIleGluLeuGluAlaGlnArg 1484
QY 886 -----CTGGCCAGAGAG 897
Db 1485 ThrLysValLeuGluLeuGluLysGlnLysAsnPhaeAspLysLysIleLeuAlaGluGlu 1504
QY 898 GAGTGGCTG-----AGACTAACAGAACTGCTGGCGAATCTGACCACTGACACCTC 951

```

```

Db 1505 LysAlaIleSerGlnGInIleAlaGInGluArgAspThrAlaGluArgGluAlaArgGlu 1524
QY 952 ACCGAGCTGCAAAATAGCTCAACTGCTACGACGAAAGAAAGGTATACATGATTAATGGA 1011
Db 1525 LysGluThrLysValLeuSerValSerArgGlnLeuAspGlnAlaPheAspLysIleGlu 1544
QY 1012 AACTTACAGAGAAATGAGAAATTTGAGCAACAGTGTGTCCAGCATGTGGAGAGTACAT 1071
Db 1545 AspLeuLysnLysArgLysThrLeuGlnAsnGluLeuAspAspLeuAlaAsnThrGln 1564
QY 1072 GAGACGATGAGCAAGAGCTTAAGCAGCTGATTAAGACACAGCCAGGCCACAGCCAG 1131
Db 1565 GlyThrAlaAspLysAsnValHisGluLeuGluLysAlaLysArgAlaLeuGluSerGln 1584
QY 1132 CTGTGCGAGCTCTCCAGCAACAGCAACCACTTCTCTGAGAGAGAGAGCTGTGCGAA 1191
Db 1585 LeuAlaGlu-----LeuLysAlaGlnAsnGluLeuLeuLysAspLeuGlnLeuThrGlu 1603
QY 1192 GAGGTGAGCCGCTGCGGACCCAGTTA 1218
Db 1604 AspAla-----LysLeuArgLeuGluVal 1611

RESULT 6
PDEL_RAT STANDARD; PRT; 4687 AA.
ID AC P30427; 008879; 008880; 008881;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Plectin 1 (PLTN) (PCN).
GN PLECL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Glial tumor;
RX MEDLINE=91268156; PubMed=2050743;
RA Wiche G., Becker B., Lubet R., Wetzler G., Castanon M.J.,
RT Hauptmann R., Stratowa C., Stewart M.;
RT "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT chain with a three-domain structure based on a central alpha-helical
RT coiled coil."
RL J. Cell Biol. 114:83-99(1991).
RN [2]
RP REVISIONS.
RC TISSUE=Glial tumor;
RX MEDLINE=96210632; PubMed=8633053;
RA Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT "Human plectin: organization of the gene, sequence analysis, and
RT chromosome localization (8q24).";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN [3]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.
RC TISSUE=Glial tumor;
RX MEDLINE=97321050; PubMed=9177781;
RA Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RA Wiche G.;
RT "Plectin transcript diversity: identification and tissue distribution
RT of variants with distinct first coding exons and rodless isoforms.";
RL Genomics 42:115-125(1997).
CC -1- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS: 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH

```



```

OY 346 GCTTATGACAGCGTGAACAGAGTTTGCAGATATCTGAGAGACCCATTGTAACCAACC 405
    ||| ||| : : : : : ||| ||| : : : : :
DB 1778 ALaasngluAlaLeuArgLeuArgLeuIn---AlaGluGluValAlaGlnGlnLysSer 1796
OY 406 AAGCCTTATCCAGTGCAGCTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 465
    ||| ||| : : : : : ||| ||| : : : : :
DB 1797 LeuAlaGluAlaAspAlaGluLysGlnLysGlnGluAlaGluAlaArgGlnAlaArgArg 1816
OY 466 GAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525
    ||| ||| : : : : : ||| ||| : : : : :
DB 1817 GlyLys-----AlaGluGluGlnAlaValArgGlnArgGlnGluAlaGluGln 1832
OY 526 GAAATATACGAGAGAGAG-----GACTATACGGGATCAAGAGAGAGAGAGAGAGAG 576
    ||| ||| : : : : : ||| ||| : : : : :
DB 1833 GluLeuGluLysGlnArgGlnLeuThrGlnGluGlnGlnGlnGlnGlnGlnGlnGln 1852
OY 577 CAGAAATATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 609
    ||| ||| : : : : : ||| ||| : : : : :
DB 1853 GlnGluLeuLeuLeuArgLeuArgAlaGluThrGlnGlnGlnGlnGlnGlnGlnGln 1872
OY 610 -----AAGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 654
    ||| ||| : : : : : ||| ||| : : : : :
DB 1873 GluGluGluLeuAlaArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1892
OY 655 ATTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 714
    ||| ||| : : : : : ||| ||| : : : : :
DB 1893 LeuGluAlaGluLeuAla-----LysValArgAlaGlnGlnGlnGlnGlnGlnGln 1905
OY 715 TATCAGCTGAAATTAACCAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 774
    ||| ||| : : : : : ||| ||| : : : : :
DB 1906 ValLeuLeu-----AlaSerLysAlaArgAlaGlnGlnGlnGlnGlnGlnGlnGln 1921
OY 775 AGAGCAAAACCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 813
    ||| ||| : : : : : ||| ||| : : : : :
DB 1922 SerGlnLysSerLysGlnArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1941
OY 814 GAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 873
    ||| ||| : : : : : ||| ||| : : : : :
DB 1942 GluAlaAlaArgLeuArgAlaLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1961
OY 874 CAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 930
    ||| ||| : : : : : ||| ||| : : : : :
DB 1962 GluAlaAlaAlaArgGlnArgAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1981
OY 931 TCTGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 987
    ||| ||| : : : : : ||| ||| : : : : :
DB 1982 IleSerGluAlaThrArgLeuLysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1999
OY 988 AGGTATACATATGATTAATTTGGAAGGTTA-----CAGACAGAGAGAGAGAGAGAG 1029
    ||| ||| : : : : : ||| ||| : : : : :
DB 2000 GluAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2018
OY 1030 GAAAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1089
    ||| ||| : : : : : ||| ||| : : : : :
DB 2019 ---ArgLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2034
OY 1090 CTAAAGCAGCTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1149
    ||| ||| : : : : : ||| ||| : : : : :
DB 2035 LeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2053
OY 1150 AAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
    ||| ||| : : : : : ||| ||| : : : : :
DB 2054 -----GluAspThrLeuArgGlnArgArgGlnValGlnGlnGlnGlnGlnGlnGln 2071

```

```

OS Dictyostellium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=96215148; PubMed=8636147;
RA Hammer J.A. III, Jung G.;
RT "The sequence of the dictyostellium myo J heavy chain gene predicts a
RT novel, dimeric, unconventional myosin with a heavy chain molecular
RT mass of 258 kDa."
RL J. Biol. Chem. 271:7120-7127(1996).
RN [2]
RP SEQUENCE OF 1-1021 FROM N.A.
RX MEDLINE=97039016; PubMed=8884597;
RA Peterson M.D., Urioste A.S., Titus M.A.;
RT "Dictyostellium discoideum myoJ: a member of a broadly defined myosin
RT V class or a class XI unconventional myosin?";
RL J. Muscle Res. Cell Motil. 17:411-424(1996).
RN [3]
RP SEQUENCE OF 182-298 FROM N.A.
RX MEDLINE=95023928; PubMed=7937787;
RA Titus M.A., Kuspa A., Loomis W.F.;
RT "Discovery of myosin genes by physical mapping in Dictyostellium.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: UA2409; AAB5186.1; -.
DR EMBL: U35322; AAB79858.1; -.
DR HSSP: P08799; 1MND.
DR DictyDb: DD01095; myoJ.
DR InterPro: IPR002710; DIL.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 2.
DR Pfam: PF00612; IQ; 6.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 2.
DR PRODOM: PD003376; DIL; 1.
DR SMART: SM00015; IQ; 3.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 3.
KW Myosin; ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
KW Coiled coil.
FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
FT DOMAIN 824 851 IQ 1.
FT DOMAIN 872 901 IQ 2.
FT DOMAIN 943 972 IQ 3.
FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
FT DOMAIN 1813 2245 TAIL.
FT NP_BIND 174 181 ATP (POTENTIAL).
FT DOMAIN 669 749 ACTIN-BINDING.
FT CONFLICT 191 191 L -> F (IN REF. 2).
FT CONFLICT 284 284 A -> T (IN REF. 2).
FT CONFLICT 291 291 G -> R (IN REF. 2).
FT CONFLICT 332 347 NKGCEFGVSDSEH -> IEMPELKYRKRS (IN
FT REF. 2).

```


FT CONFLICT 550 550 N -> K (IN REF. 2).
 FT CONFLICT 865 866 HH -> QQ (IN REF. 2).
 SQ SEQUENCE 2245 AA; 258478 MW; 615E5E1D1A845BE CRC64;

 Alignment Scores:
 Pred. No.: 1,49e-06 Length: 2245
 Score: 245.00 Matches: 108
 Percent Similarity: 37.94% Conservative: 87
 Best Local Similarity: 21.01% Mismatches: 195
 Query Match: 8.97% Gaps: 124
 DB: 1 Indels: 12

 US-09-502-945-1 (1-1552) x MYSJ_DICDI (1-2245)

 QY 4 CTGATGATCCGAGAACTAAACCTTATGAGAAAGTGAATGGAATCC 63
 DB 1064 LeuaspHisSerLysLeuAsnLysLysLeuGluLysAspLeuSerAspGlnHisAsp 1083
 QY 64 CAATTGAAGTTTGGAGAACGACTGATGATATCAGAACCTTGAGACTTTAA 123
 DB 1084 SerIleGluLysLeuGlnSerGlnPheAsnGlnThrGlnGlnGlnPheLys 1103
 QY 124 GAGCACTAAAGCATTAAGAAATTTCTGCTGCT 159
 DB 1104 GlnGlnSerGlnGlnLeuSerSerLysLeuSerLysThrGlnGlnLeuAspPheAsn 1123
 QY 160 -----AATCTGTAAACCGTGT 177
 DB 1124 LysGlnGlnPheAspArgLeuSerGlnGlnLysAspThrAspAsnThrAsnGlnLeu 1143
 QY 178 GGTGCTCTTTGAAATGTGCT-----CAGCATGAAGCTGTCTTCC 222
 DB 1144 GluIleGlnGlnLeuLysLysAlaAsnSerThrLeuGlnGlnLysAspPheSerLeuSer 1163
 QY 223 CAAACCCACTAATGTTTCATATGACCATCGAA-----AGACTGTT 267
 DB 1164 GlyIleArgAspAsnLeuGlnArgGlnValLeuGlnLeuArgAspGluAsnGlnLeuIle 1183
 QY 268 AAAAAGAGATGACTGATGCTGCTGCACTACTTCCGTAAGAGAGCTTGCA----- 321
 DB 1184 LysGlnGlnLeuAspSerLeuGlnGlnSerGlnPheGlnSerGlnValAlaLeu 1203
 QY 322 GATACGACGAAAGAGCAAGAGTCTTATGACAGGTGAAGCAAGTTTCAATATCT 381
 DB 1204 GluLysGlnGlnLeuGlnGlnLeuValGlnGlnSerGlnGlnLeuIleLysLeuSer 1223
 QY 382 GAGCAAGCCAAATTTGAAAAACCAAGCCTTAAATCCAGTGTGACCAAGTGAAGAG 441
 DB 1224 SerGlnLysLeuGlnSerGlnGlnGlnLysGlnIleAsnGlnLeuGlnLeu 1243
 QY 442 CTGAGAGAGCGAGGCGGAGCGACTTGAAAGAAAGACTGCACTTCAGCAAGAAAGGCGC 501
 DB 1244 LeuThrAspHisLysSerLysLeuGlnIleGlnLeu-----GlnLeuThrGlnGlnSer 1261
 QY 502 ATTGAGAAAGCATGATGATGAAAGAAATTAACGAAAGAAAGGAGCATGAGTCAAG 561
 DB 1262 AsnGlnLys-----IleLysLysLeuLysGlnLysLeuGlnLysGlnLysGlnLys 1279
 QY 562 ATGTTGATCTTGTCTCAGATATTTGCCCACTGAGGCCCAAGTGAAGAGTTACAAG 621
 DB 1280 -----LysGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1291
 QY 622 GAAAAGATTTACGCTATTATCACTGAGAGAAATTCAAACCGACTGCTTCGGGAA 681
 DB 1292 SerLysGlnSerValGlnAspGlnLysAsnSerLeuIleThrGlnLeuThrThrValLys 1311
 QY 682 ATGATGATCACAAGAGTGTGAGAAATGCGCTATCAGATGCAATTAACAACATGAG 741
 DB 1312 PheGlnSerThrGlnValSerThrAsnValSerHisGlnLysGlnLysIleThrThrLeu 1331
 QY 742 AAGATGAGGAGCAAGAAAGACACAGAGATTCAAGCAAAAACATACAGGAGATCTGAA 801
 DB 1332 LysSerThrIleGlnGlnLeuAsnLysSerIle---GlyLysLeuGlnAlaGlnGlnLys 1350

QY 802 ATTAAGATCAGAAATAGAGAAATTGGAATAGACTGATGAAGCAACACAC---- 858
 DB 1351 AsnLysAspAspGlnIleArgLysIleGlnPheGlnLeuAsnAspGlnLysGlnPhe 1370
 QY 858 ----- 858
 DB 1371 ThrArgGlnThrLysGlnPheSerAspLeuGlnSerGlnGlnSerIleAspArgProLys 1390
 QY 859 -----TTGGAA 864
 DB 1391 SerGlnIleThrIleHisSerLeuGlnLysArgThrAsnGlnThrLeuLysSerAspPheGln 1410
 QY 865 CAGGACGACGACGAGGCGACCCCTGGCCACAGAGAGCTC----- 903
 DB 1411 ArgValGlnGlnSerLeuLysGlnGlnGlnLysLysArgAspGlnThrLysAspThrIle 1430
 QY 904 -----CTGAGACTAACAAGACTGTGGCGAATCTGAGCACCAC 942
 DB 1431 AsnArgLeuGlnAsnGlnValLysGlnLeuThrGlnLeuLysGlnLysArgPheGlnAsnGln 1450
 QY 943 CTGCACCTCACACAGATCTGAAATAGCTCACTCACTCAAGAAAAAGTATACATATCAT 1002
 DB 1451 PhePheValAlaLysGlnGlnAsnSerAsnGlnThrGlnGlnSerValThrLeuLysGln 1470
 QY 1003 AAATTGGGAAAGTTACAGAGAAAGAAATGAGAAATGAGGAGCAAGTGTCCAGCATCGG 1062
 DB 1471 ValThrThrGlnMetGlnGlnAsnGlnSerArgIleGlnLysGlnLeu----- 1486
 QY 1063 AGAGTACATGAGAGCATGACGATGACCAAGGCTAAGGCGTGAAGCAGCAGCAGCCACA 1122
 DB 1487 -----GlnGlnLysLysGlnHisIleThrArgIleAspAspGlnArgAspGlnLeu 1503
 QY 1123 GCCACAGCTGTGTCACCTCTCAGCAGCAAGCAAGCTTCTCTGAGAGGACAGGC 1182
 DB 1504 LysLysGlnLeuThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1523
 QY 1183 CTGTCGGAAGAGGTGACGCGGTGCGACCCAGTACCAGCATCCACAAATCTGATTCG 1242
 DB 1524 AlaGlnAsnGlnLeuGlnLysArgLys----- 1533
 QY 1243 TGACCTGATGAGAACAGATGAATTAATTAATTAACAAGAG 1284
 DB 1534 -----LysGlnLeuLysThrLysGln 1540

 RESULT 8
 MYH3_CHICK STANDARD; PRT; 1940 AA.
 ID MYH3_CHICK AC P02565;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin heavy chain, fast skeletal muscle, embryonic.
 GN MYH3.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87194881; PubMed=3571266;
 RA Molina M.I., Kropp K.E., Gulick J., Robbins J.;
 RT "The sequence of an embryonic myosin heavy chain gene and isolation
 of its corresponding cDNA."
 RL J. Biol. Chem. 262:6478-6488(1987).
 RP [2]
 RP SEQUENCE OF 1502-1940 FROM N.A.
 RX MEDLINE=83161144; PubMed=6833296;
 RA Kavinsky C.J., Umeda P.K., Sinha A.M., Elzinga M., Tong S.W., Zak R.,
 RA Jakovlev S., Rabinowitz M.;
 RT "Cloned mRNA sequences for two types of embryonic myosin heavy chains
 from chick skeletal muscle. I. DNA and derived amino acid sequence of

Db 1574 LysSerAspIleAspArgIleAlaGluLysAspGluIleAspGluLysArg 1593
 QY 835 GAA-----CTGGATGAAGCAACACTGTGAACAGGACAGCAGAGCA 882
 Db 1594 AsnHisLeuArgValValAspSerMetClnSerThrLeuAspAlaGluIleArg----- 1611
 QY 883 GCCCTGGCCAGAGAGACTGCTGAGACTTAACAGAACTG-----CTGGGCCGA 930
 Db 1612 -----SerArgAsnGluAlaLeuArgLeuLysLysMetGluGlyAspLeuAsnGlu 1629
 QY 931 TCTGAGCACCACACTGCACCTCACCAGATCTGAATAGCTCACTCAGTCAAGAAAAAG 990
 Db 1630 IleGluIleGlnLeuSerHisAlaAsnArgGlnAlaAlaGluAlaGlnLysAsnLeuArg 1649
 QY 991 TATACATATGATTAATTTGGGAAGTTACAG-----AGAAGAAAT 1029
 Db 1650 AsnThrGlnGlyValLeuLeuLysAspThrGlnIleHisLeuAspAlaLeuArgSerCln 1669
 QY 1030 GAAGAATTGGAGAACTGTGTCACAGTGGAGA-----GTA 1068
 Db 1670 GluAspLeuLysGluGlnValAlaMetValGluArgArgAlaAsnLeuLeuGlnAlaGlu 1689
 QY 1069 CATAGAGCATGATGAAGCAAGCTAAGGAGCTGATTAAGCAGCCAGCCAGCCAG 1128
 Db 1690 IleGluGluLeuArgAlaAlaLeuGlnGlnThrGluArgSerArgLysValAlaGluGln 1709
 QY 1129 CAGCTG-----GTGCACTGCTCCAGCAGCAAGCAAGCACTCTCTCTG 1170
 Db 1710 GluLeuLeuAspAlaSerGluArgValGlnLeuLeuHisThrGlnAsnThrSerLeuIle 1729
 QY 1171 GAG---AGCGAGAGCTGTCTCGAAGAGGTGGACCGGCTGCGGACCACTTA 1218
 Db 1730 AsnThrLysLysLysLeuGluSerAspLeuSerGlnIleGlnSerGluMet 1746
 RESULT 9
 ID PLE1_HUMAN STANDARD; PRT; 4684 AA.
 AC Q15149; Q16640; Q15148;
 DT 16-OCT-2001 (Ref. 40, Created)
 DT 16-OCT-2001 (Ref. 40, Last sequence update)
 DT 16-OCT-2001 (Ref. 40, Last annotation update)
 DE Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).
 GN PLECL.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Placenta;
 RX MEDLINE=96210632; PubMed=8633055;
 RA Liu C.-G., Maerker C., Castanon M.J., Hauptmann R., Wiche G.;
 RT "Human plectin: organization of the gene, sequence analysis, and
 RL chromosome localization (8q24)."
 RL Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
 RP [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.
 RX MEDLINE=96312447; PubMed=8696233;
 RA McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,
 RA Bullrich F., Burgesson R.E., Amano S., Hudson D.L., Owaribe K.,
 RA McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,
 RA Uitto J.;
 RT "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:
 RT cDNA cloning and genomic organization."
 RT Genes Dev. 10:1724-1735(1996).
 RN [3]
 RP VARIANT MD-EBS 1003-GLN-ALA-1005 DEL.
 RX MEDLINE=97049959; PubMed=8894687;
 RA Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,
 RA Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.;
 RT "Homozygous deletion mutations in the plectin gene (PLECL) in patients
 RT with epidermolysis bullosa simplex associated with late-onset
 RT muscular dystrophy."

RL Hum. Mol. Genet. 5:1539-1546(1996).
 RP [4]
 RN VARIANT MD-EBS LEU-429 INS.
 RX MEDLINE=21090821; PubMed=11159198;
 RA Bauer J.W., Rouan F., Kofler B., Reznicek G.A., Kornacker I.,
 RA Muss W., Hamelner R., Klausegger A., Huber A., Pohl-Gubo G.,
 RA Wiche G., Uitto J., Hintner H.;
 RT "A compound heterozygous one amino-acid insertion/nonsense mutation in
 RT the plectin gene causes epidermolysis bullosa simplex with plectin
 RT deficiency."
 RL Am. J. Pathol. 158:617-625(2001).
 CC -1- FUNCTION: INTERLINS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
 CC MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
 CC HEMIDESMOSOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO
 CC MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE
 CC CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE
 CC FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.
 CC -1- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN
 CC MUSCLE, HEART, PLACENTA AND SPINAL CORD.
 CC -1- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
 CC VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
 CC AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
 CC -1- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
 CC INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
 CC -1- DISEASE: DEFECTS IN PLECL ARE THE CAUSE OF EPIDESMOLYSIS BULLOSA
 CC WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-KD), AN AUTOSOMAL RECESSIVE
 CC DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL
 CC OF THE HEMIDESMOSOME AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 PLECTIN REPEATS.
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: Z54367; CA91196.1; -;
 DR EMBL: U53204; AAB05427.1; -;
 DR EMBL: U63610; AAB05428.1; -;
 DR EMBL: U63609; AAB05428.1; JOINED.
 DR EMBL: X97053; CAA65765.1; -;
 DR HSSP: Q01082; IBKR
 DR Genew: HGNC:9069; PLECL.
 DR MIM: 601282; -;
 DR MIM: 226670; -;
 DR InterPro: IPR001589; Actbind_actinin.
 DR InterPro: IPR001715; Calponin-like.
 DR InterPro: IPR001101; Plectin_repeat.
 DR InterPro: IPR005326; S10_Plectin.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00307; CH; 2.
 DR Pfam: PF00681; Plectin; 19.
 DR Pfam: PF03501; S10_Plectin; 1.
 DR SMART: SM00033; CH; 2.
 DR SMART: SM00250; PLEC; 33.
 DR SMART: SM00150; SPEC; 5.
 DR PROSITE: PS00019; ACTININ_1; FALSE_NEG.
 DR PROSITE: PS00020; ACTININ_2; FALSE_NEG.
 DR PROSITE: PS50021; CH; 2.
 DR Coiled coil: Repeat: structural protein: Cytoskeleton; Actin-binding;
 KW Phosphorylation; Alternative splicing; Epidermolysis bullosa;
 KW Disease mutation.
 FT DOMAIN 1 1470 GLOBULAR 1.
 FT DOMAIN 1471 2755 CENTRAL FIBROUS ROD DOMAIN.

Db 1656 SerATgValtysAlaGluAlaGluAlaAlaArgGluTysGlnArgAlaLeuGlnAlaLeu 1675
 Qy 649 GAGGAAATTCAGAGCCAGCTGGCTTCTCGGGAAATGATGTGCACAAAGTGTGTGAGAA 708
 Db 1676 Gt 1695
 Qy 709 ATGCCTATTCAGCTGATAT-----AAACCAACATGAGACGATGAGCAGCAAAAG 759
 Db 1696 ArgAlaArgGlnValGlnValAlaLeuGlnTrrAlaGlnArgSerAlaGlnAlaGlnLeu 1715
 Qy 760 GAGCAGAGA-----GAGTTTCAGAGCAAAACTTAACAGGATCTTGAATTTAAAGATCAG 813
 Db 1716 GlnSerLysArgAlaSerPheAlaGluTyrThr-----AlaGlnLeuGlnArgSerLeuGln 1734
 Qy 814 GAAATAGAGAAATTCAGATGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 Db 1735 Gt 1754
 Qy 874 CAGAGAGGAGCCCTGCGCAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 924
 Db 1755 AlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGluAlaGlu 1774
 Qy 925 GGCAGATTCAGAGCCAGCTGACCTCAGCAGATCTGAAATA-----CTGCTG 966
 Db 1775 AlaAngLualAlaLeuArgLeuArgLeuGlnAlaGluGlnValLeuGlnGlnLysSerLeu 1794
 Qy 967 GCTCACTCACTGACGAGAAAAAGTATACATATGATTAATTTGGAAAGTTTACAGAGAGA 1026
 Db 1795 AlaGlnAlaGluAlaGluAlaGluGln-----LysGlnGluAlaGluAlaGluAlaArgArg 1812
 Qy 1027 AATGAAGAATTGGAGAGACAGTGTGTCCAGATGAGGAGAGATCATAGAGCGATGAAGCAA 1086
 Db 1813 ArgGlyLysAlaGlnGlnGlnGlnAlaValArgGlnArgGlnArgGlnGlnGlnGlnGln 1832
 Qy 1087 AGGCTAAGCAGACGATGATTAAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1131
 Db 1833 LysGlnArgGlnLeu-----AlaGlnGlyThrAlaGlnGlnArgLeuAlaGlnAlaGln 1849
 Qy 1132 -----CTGGTGCAGACTCTCTCAGCAGAGCAAGACGATTTCTCTCTGAGAGGAGGAGCCTG 1185
 Db 1850 Gt 1869
 Qy 1186 TCGAAGAGGTGAGCGGCTGCGAGCCAGTACCAGTACCAGTACCAGTACCAGTACCAGTACCAGT 1233
 Db 1870 Gt 1885
 RESULT 10
 MYH9_HUMAN STANDARD: PRT: 1960 AA.
 ID AC P35579; 060805;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Myosin heavy chain, nonmuscle type A (Cellular myosin heavy chain,
 type A) (Nonmuscle myosin heavy chain-A) (NMHC-A).
 GN MYH9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057165; PubMed=10591208;
 RA Dunham I., Hunt A.R., Collins J.E., Bruskiewicz R., Beare D.M.,
 RA Clump M., Smink L.J., Alnsough R., Almeida J.P., Babbage A.K.,
 RA Bauguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
 RA Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,
 RA Burdill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
 RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
 RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
 RA Dhani P.D., Dockree C., Dodsorth S.J., Durbin R.M., Ellington A.G.,
 RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,

RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
 RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
 RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
 RA Laird G.K., Langford C.F., Leverisa N.A., Lloyd C., Lloyd D.M.,
 RA Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
 RA Mcclay N., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
 RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
 RA Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
 RA Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,
 RA Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
 RA Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
 RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
 RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
 RA Mioshlma S., Kawasaki K., Sasaki T., Aakawa S., Kudoh J.,
 RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
 RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
 RA Dorfman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
 RA Lewis J., Lewis S., Lin S.-P., Loh P., Malat E., Nguyen T., Pan H.,
 RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
 RA Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
 RA Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Mink P.,
 RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
 RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
 RA Hinds K., Kemp K., Latreille P., Layman D., Ozerisky P., Rohlfing T.,
 RA Scheet P., Walker C., Wamsley A., Wohlmann P., Pepin K., Nelson J.,
 RA Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
 RA Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budari M.L.,
 RA Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
 RA Kim U.J., Shizuya H., Simon M.I., Dunamski J.P., Peyrad M., Kedra D.,
 RA Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,
 RA Wilkison P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
 RA Tliahun Y., Wright H.;
 RA "The DNA sequence of human chromosome 22.";
 RL Nature 402:489-495(1999).
 [2]
 RP SEQUENCE OF 1-1337 FROM N.A.
 RX MEDLINE=92003925; PubMed=1912569;
 RA Tootcher L.E., Gonzalez D.A., Tung N., Lemons R.S., Le Beau M.M.,
 RA Arnout M.A., Clayton L.K., Tenen D.G.,
 RA "Cellular myosin heavy chain in human leukocytes: Isolation of 5'
 RT cDNA clones, characterization of the protein, chromosomal
 RT localization, and upregulation during myeloid differentiation.";
 RL Blood 78:1826-1833(1991).
 [3]
 RP SEQUENCE OF 1-715 FROM N.A.
 RX MEDLINE=91316803; PubMed=1860190;
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Gdula D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located
 RT on different chromosomes.";
 RL Circ. Res. 69:530-539(1991).
 [4]
 RP SEQUENCE OF 714-1960 FROM N.A.
 RX MEDLINE=90138958; PubMed=1967836;
 RA Saer C.G., Myers J.C., Shows T.B., Leinwand L.A.;
 RT "Human nonmuscle myosin heavy chain mRNA: generation of diversity
 RT through alternative polyadenylation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1164-1168(1990).
 [5]
 RP VARIANT DFNA17 HIS-705.
 RX MEDLINE=20489856; PubMed=11023810;
 RA Lalwani A.K., Goldstein J.A., Kelley M.J., Luxford W., Castelein C.M.,
 RA Mhatare A.N.;
 RT "Human nonsyndromic hereditary deafness DFNA17 is due to a mutation in
 RT nonmuscle myosin MYH9.";
 RL Am. J. Hum. Genet. 67:1121-1128(2000).
 [6]
 RP VARIANTS MHA/FTNS/SBS K-93; C-702; C-1165; H-1424 AND K-1841.
 RX MEDLINE=20428192; PubMed=10973259;
 RA Seri M., Cusano M., Gaugiarosa S., Caridi G., Bordo D., Lo Nigro C.,
 RA Ghigeri G.M., Ravazzolo R., Savino M., Del Vecchio M., d'Apolito M.,
 RA Iolascon A., Zelan L.L., Savola A., Baldini C.L., Norris P.,
 RA Magrini U., Belletti S., Heath K.E., Babcock M., Gluckman M.J.,

RA Allprandis E., Bizaro N., Desnick R.J., Martignetti J.A.;
 RT "Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and
 RT Sebastian syndromes.";
 RL Nat. Genet. 26:103-105(2000).
 RN [71]
 RP VARIANTS MHA ILE-1155 AND LYS-1841.
 RX MEDLINE-20428193; PubMed-10973260;
 RA Kelley M.J., Javien W., Ortel T.L., Korczak J.F.;
 RT Mutation of MYH9, encoding non-muscle myosin heavy chain A, in
 RT May-Hegglin anomaly.";
 RL Nat. Genet. 26:106-108(2000).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CC CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CC CAPING.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CC CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 CC REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF MAY-HEGGLIN ANOMALY
 CC (MHA), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC (FTNS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC WITH ADDITIONAL ALPOPT-LIKE CLINICAL FEATURES OF SENSORINEURAL
 CC DEAFNESS, CATARACTS AND NEPHRITIS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF SEBASTIAN SYNDROME
 CC (SBS), AN AUTOSOMAL DOMINANT MACROTHROMBOCYTOPENIA CHARACTERIZED
 CC BY THROMBOCYTOPENIA, GIANT PLATELETS AND LEUCOCYTE INCLUSIONS.
 CC -1- DISEASE: DEFECTS IN MYH9 ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DEFA17) WHICH IS
 CC CHARACTERIZED BY PROGRESSIVE HEARING IMPAIRMENT AND
 CC COCHLEOSACCULAR DEGENERATION.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 TO DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: 282215; CAB05105.1; -
 DR EMBL: M81105; AAA59888.1; -
 DR EMBL: M69180; AAA61765.1; -
 DR EMBL: M31013; AAA36349.1; -
 DR HSSP: P10587; 1BR2.
 DR GeneW: HGNC:7579; MYH9.
 DR MIM: 160775; -
 DR MIM: 153640; -
 DR MIM: 155100; -
 DR MIM: 603622; -
 DR MIM: 605249; -
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF01576; Myosin_tail; 1.
 DR Pfam: PF02736; Myosin_N; 1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR PRODOM: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ; 1.
 DR Myosin: ATP-binding; Calmodulin-binding; Actin-binding;
 KW Coiled coil; Alkylation; Multigene family; Disease mutation;

KW Deafness. 1 778 MYOSIN HEAD-LIKE.
 FT DOMAIN 779 808 IQ.
 FT DOMAIN 779 808 COILED COIL (POTENTIAL).
 FT DOMAIN 837 1926 ATP (POTENTIAL).
 FT NP_BIND 174 181 ACTIN-BINDING.
 FT DOMAIN 654 676 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 694 694 ALKYLATION (SH-2) (POTENTIAL).
 FT MOD_RES 704 704 N -> K (IN MHA).
 FT VARIANT 93 93 /FTid=VAR.010791.
 FT VARIANT 702 702 R -> C (IN FTNS).
 FT VARIANT 702 702 /FTid=VAR.010792.
 FT VARIANT 705 705 R -> H (IN DENA17).
 FT VARIANT 705 705 /FTid=VAR.010793.
 FT VARIANT 1155 1155 T -> I (IN MHA).
 FT VARIANT 1155 1155 /FTid=VAR.010794.
 FT VARIANT 1165 1165 R -> C (IN SBS).
 FT VARIANT 1424 1424 /FTid=VAR.010795.
 FT VARIANT 1424 1424 D -> H (IN FTNS).
 FT VARIANT 1841 1841 /FTid=VAR.010796.
 FT VARIANT 1841 1841 E -> K (IN MHA).
 FT CONFLICT 53 55 /FTid=VAR.010797.
 FT CONFLICT 660 660 EAI -> S (IN REF. 3).
 FT CONFLICT 869 869 T -> S (IN REF. 3).
 FT CONFLICT 931 931 T -> M (IN REF. 4).
 FT CONFLICT 1240 1241 C -> Y (IN REF. 4).
 FT CONFLICT 1241 1241 KG -> GR (IN REF. 4).
 FT CONFLICT 1350 1350 E -> EE (IN REF. 2).
 FT CONFLICT 1764 1764 T -> A (IN REF. 2).
 FT CONFLICT 1771 1771 S -> G (IN REF. 2).
 Alignment Scores:
 Pred. No.: 1.95e-06 Length: 1960
 Score: 243.00 Matches: 106
 Percent Similarity: 40.87% Conservative: 91
 Best local Similarity: 21.99% Mismatches: 161
 Query Match: 8.89% Indels: 124
 DB: 1 Gaps: 17
 US-09-502-945-1 (1-1552) x MYH9_HUMAN (1-1960)
 QY 7 GATGCAATCCGAGCAATTAAGTACTTATGAGAAAG----- 45
 Db 1409 AspIysLeuGIuIuIyStHrIyStHrArGIeudInIuGluLeuAspPleuValAsp 1428
 QY 46 -----TGTGAATTGAGCAATCCCAATTGAAGTTT----- 75
 Db 1429 LeuAspHISGInArGIInserAlaCyAsnLeuIuIySgInIuIySphAspGIu 1448
 QY 76 -----TTGAGCAAGCACTTACGTGAATATACAGAACTTGTGAAGATCTTAA 123
 Db 1449 LeuLeuAlaGIuIuIyStHrIleSerAlaIyStyAlaGIuIuIyAspArgAlaGIu 1468
 QY 124 GAGCAACTAAAGCATTAAGATTTCTTGCGCTCTAATCTTAACCGGTGGGCT 183
 Db 1469 AlaGIuAlaArgGIuIySgIu-----ThIyAla 1478
 QY 184 CTTTGTGGAATGTCTGCTGAGCATGAGCTGTCTTCCCAACCCTACTAATGTCAT 243
 Db 1479 LeuSerLeuAlaArgAlaIaLeuGIuIuIyAlaMetGIuIuIyAlaGIu----- 1494
 QY 244 ATGCAGACATCGAAGACGTGGTTAA-----GAAGAGATGACTTGATGTCT 291
 Db 1495 -----LeuGIuArgLeuAsnIySgInPheArgThrGIuMetGIuAspLeuMetSer 1511
 QY 292 GCACTAGTTCCTGAAGAGACAGCTTGCGCATGATGCGAGAGAGAGAGAGAGCTTAT 351
 Db 1512 SerIyAspAspValGIyIySgIuIySgIuIySgIuIySgIuIySgIuIySgIuIy 1531
 QY 352 GAACAGGTGAACCACTT-----TTGCAATATCT 381
 Db 1532 GInGIuValGIuIuIyMetIyStHrGIuLeuGIuIuIySgIuIuIySgIuIuIyAlaThr 1551


```

US-09-502-945-1 (1-1552) x CENE_HUMAN (1-2663)
OY 1 CTCTGTGATCATCCGAGAGCTAAACCTACTTATGAGAAAAGTGAATTGAGAA 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 722 IIEThAspleuclnlysgluleasnlysluvalgluInuInglualaleuarglu 741
OY 61 -----TCCCAATTGAAGTTTGTGAGAACGACTAGCTGAATATCAGAGA 105
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 742 GluValIIEleuLeuSerGluLeuLysSerLeuProSerGluValIGluArgLeuArgLys 761
OY 106 ACTTGTGAGAGATCTTAAAGCACTA-----AAGCATTAAGAATT 147
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 762 GluIIEGlnAspLysSerGluGluLeuHisIIEThrSerGluLysAspLysLeuPhe 781
OY 148 CTCTGTGCTCTAATCTTAAACCGTGTGCTGTCTTGTGAATGCTGCAGCAT 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 782 SerGluValValHisLysGluSerArgValGlnGlyLeuGluGluIIEGlyLysThr 801
OY 208 GAAGCTGTCTTCCCAACCCACTACTAAT----- 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 802 LysAspAspLeuAlaThrThrGlnSerAsnTyrLysSerThrAspGlnGluPheGlnAsn 821
OY 238 -----GTCAATG----- 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 822 PheLysThrLeuHisMetAspPheGluGlnLysTyrLysMetValLeuGluGluAsnGlu 841
OY 247 -----CAGACCATCGAAGACAGCTGTAAAGAAAAGATGACTGTCTGCACCTA 297
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 842 ArgMetAsnGlnGluIIEValAsnLeuSerLysGluIIEGlnLysPheAspSerLeu 861
OY 298 GTTCCGTAAAGAGAGAGCTTGCA-----GATACGACGAGAAAGAGAGCAAGT 345
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 862 GlYAlaLeuLysThrGluLeuSerTyrLysThrGlnGluLeuGlnGluLysThrArgGlu 881
OY 346 GCTTATGAACAGCTGAAACACTTTTGCATAATCTGAGCA----- 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 882 ValGlnGlnArgLeuAsnGluMetGluGlnLeuLysGluGlnLeuGluAsnArgAspSer 901
OY 388 -----GCCAATTTTGAAAAAACCAAGGCTTTAATCCAGTGCAGTGAAGAGAGAG 441
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 902 ProLeuGlnThrValGlnArgGluLysThrLeuIIE-----ThrGluLysLeuGlnGlnThr 920
OY 442 CTGAGAGAGCGAGCGAGCACTTGAAAAAGAA-----CTTCATCTCAGCAAGAG 492
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 921 LeuGluL-----GluValLysThrLeuThrGlnGluLysAspAspLeuLysGlnLeuGlnGlu 939
OY 493 AAAAGGCGCATTCAGAAACATGATGAAAAAGAAATACGAAAGAAAGGAGTACATG 552
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 940 SerLeuGlnIIEGlnArgAspGlnLeuLysSerAspIIEHisAsp----- 954
OY 553 GGATCAAAAGATTTGATCTGTCTCAAAATATT-----GCCCAACTGAGAGCCAG 603
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 955 -----ThValAsnMetAsnIIEAspThrGlnGlnGlnLeuAlaArgAsnAla 969
OY 604 GTGGAAGAGTTTACAAAGAAAAGATTTCACTAATTATCACTGAGAAATTCAAAGC 663
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 970 LeuGlnSerLeu-----LysGlnHisGlnGlnIIEAsnThrLeuLysSerLysIIEser 988
OY 664 CAGCTGCTTCCTCGGGAATGATGTCACAAAGCTGTGTGGAATAATGCC-----TAT 717
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 989 GluGlnValSerArgAsnLeuHisMetGluGluAsnThrGlyLysThrLysAspGluPhe 1008
OY 718 CAGCTGAATTAACCAACATGAGAGATGAGAGAAAGAGAAAGAGACAGAGATTGACA 777
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1009 GlnGlnLysMetValGlyLysLysGln-----AspLeuGln 1022
OY 778 GCAAAAAATCAACAGGAGCTT-----GAAATTAAAGATCAGAAATA-----GAG 822
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1023 AlalysAsnThrGlnThrLeuThrAlaAspValLysAspAsnGlnIIEIIEGlnGln 1042
OY 823 AAATTAGATAGACTGTAAGCAACAACACTTGGACAGAGACGACGAGAAAGCA 882
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 1043 ArgLysIIEPheSerLeuIIEGlnGluLysAsnGlnLeu-----GlnGlnMetLeuGluSer 1061
OY 883 GCCCTGGCCAGAGAGAGAGCTCTGAGACTA-----ACAGAACTGCTGGC 927
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1062 ValIIElaGluLysGlnGlnLeuLysThrAspLeuLysGlnLysnIIEGlnMetThrIIE 1081
OY 928 GAATCTGAGCAGCAACATGCACCTCACAGATCTGAATFACCTCAA-----CTC 975
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1082 GluAsnGlnGlnGluLeuArgLeuGlnLysAspGluLeuLysLysGlnGlnIIEVal 1101
OY 976 AGTCAGAAAAAAGCTTATCATATGATAAATTGGGAAGTTACAGAGAACAAATGAAGA 1035
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1102 AlaGlnGlnLysAsnHisAlaIIELysGlnGlyLeuLeuSerArgThrCysAspArg 1121
OY 1036 TTGAGAGAACAGTGTGTCCAGATGGAGAGATGACATGACAGATGAAGCAAGAGCTAAG 1095
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1122 LeuAlaGlu-----ValGlnGlnLysLeuLysGlnLysSerGln 1134
OY 1096 CAGCTGATTAAGCAGACAGCCAGCCAGCCAGCAGCTGTGACACTCTCAGCAAGCAG 1155
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1135 GlnLeuGlnGlnLysGlnGln-----GlnLeuLeuAsnValGln 1147
OY 1156 AACGAGCTTCTCTGAGAGAGCAGAGAGCTCTCGGAAGAGTGGACCGCTCGGAGCCAG 1215
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1148 GluGln-----MetSerGluMetGlnLysLysIIEAsnGlnIIEGlnAsnLeuLysAsnGln 1166
OY 1216 TTA 1218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1167 Leu 1167

RESULT 12
PEPL_MOUSE
ID PEPL_MOUSE STANDARD: PRT; 1755 AA.
AC Q9R269: Q9JUZ7; Q9CUP1; Q70231;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Periplakin.
GN PPL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA de Vries-Smits A.M.M., Waldmann V., Burgering B.M.T.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/Sv;
RA RYOY Y.W., Li K., Aho S., Cho B.H., Klement J.F., Uitto J.;
RT "Mouse periplakin: genomic cloning and gene targeting.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1643-1755 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gasterland T., Gliss C., King B., Kochila H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudl F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
RA Blake J., Boileau D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

```


RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA "Functional annotation of a full-length mouse cDNA collection.",
 RT Nature 409:685-690(2001).
 RL (4)
 RN SEQUENCE OF 1647-1755 FROM N.A.
 RP STRAIN=C57BL/6J;
 RC MEDLINE=98190524; PubMed=9521878;
 RX Aho S., McLean W.H.T., Li K., Uitto J.;
 RA "cDNA cloning, mRNA expression, and chromosomal mapping of human and
 RT mouse periplakin genes.",
 RL Genomics 48:242-247(1998).
 CC -1- FUNCTION: COMPONENT OF THE CORNIFIED ENVELOPE OF KERATINOCYTES.
 CC MAY LINK THE CORNIFIED ENVELOPE TO DESMOSOMES AND INTERMEDIATE
 CC FILAMENTS (BY SIMILARITY).
 CC -1- SUBUNIT: MAY FORM A HOMODIMER OR A HETERODIMER WITH EVPL (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: ASSOCIATED WITH DESMOSOMES AND INTERMEDIATE
 CC FILAMENTS (BY SIMILARITY).
 CC -1- SIMILARITY: CONTAINS 2 PLECTIN REPEATS.
 CC -1- SIMILARITY: BELONGS TO THE PLEKIN OR CYTOLINKER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AF126834; AAD20642.1; -
 DR EMBL: AF116523; AAF29436.1; -
 DR EMBL: AF116519; AAF29436.1; JOINED.
 DR EMBL: AF116520; AAF29436.1; JOINED.
 DR EMBL: AF116521; AAF29436.1; JOINED.
 DR EMBL: AF116522; AAF29436.1; JOINED.
 DR EMBL: AK014700; BAB29510.1; -
 DR EMBL: AF013715; AAC40068.1; -
 DR MGD: MGI:1194896; Ppl.
 DR InterPro: IPR001101; Plectin_repeat.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00681; Plectin_1.
 DR SMART: SM00250; PLEC: 2.
 DR SMART: SM00150; SPEC: 3.
 DR Keratinization; Repeat; Coiled coil; Cytoskeleton; Structural protein.
 KM
 FT DOMAIN 16 125
 FT DOMAIN 182 387
 FT REPEAT 214 315
 FT REPEAT 321 483
 FT REPEAT 503 610
 FT DOMAIN 611 819
 FT DOMAIN 883 1644
 FT REPEAT 1650 1684
 FT REPEAT 1699 1734
 FT REPEAT 166 166
 FT REPEAT 592 592
 FT REPEAT 648 648
 FT REPEAT 671 672
 FT REPEAT 689 721
 FT
 FT CONFLICT 983 983
 FT CONFLICT 1325 1325
 FT CONFLICT 1344 1345
 FT
 SQ SEQUENCE 1755 AA; 204003 MW; 3FEA343086E4CB8F CRC64;
 Alignment Scores:
 Pred. No.: 2.38e-06
 Score: 241.50
 Percent Similarity: 46.04%
 Best Local Similarity: 22.47%
 Length: 1755
 Matches: 102
 Conservative: 107
 Mismatches: 162

Query Match: 8.84% Indels: 83
 DB: 1 Gaps: 17
 US-09-502-945-1 (1-1552) x PEPL_MOUSE (1-1755)
 QY 16 GAGAGCTAAACTTACTATGAGAAAGTGT---GAAATT----- 54
 DB 1162 GUAAnAlAlAlValValAlGInGluValAlArgValIleValArgProAspProLys 1181
 QY 55 GAGGAATCCCAATTGAGTTTGGAGAACGACTTATGCTGAATATACAGAACTTGTA 114
 DB 1182 AlAGuSerGIuValAlAlAsnLeuArgLeuLeuValGInGInGInArgLysPheArg 1201
 QY 115 GATCTAAAGACACACTAAGACATAAAGATTCTTCGGCTGCTTAATCTGTAACGCT 174
 DB 1202 GYAlAGInGluGInLeuLeuLysSerTyrGInSerGIuLeuAla---LeuArgAsnArg 1220
 QY 175 GTTGGTGTCTTTTGTGAATGTCTCAGCATGAGCTGTTCCCAAC----- 228
 DB 1221 GYProGInValGInValLysGInValThrLysGInValIleLysTyrThrAspPro 1240
 QY 229 CATCTAATGTTCAATATGACACATCCAA-----AGACTGGTT 267
 DB 1241 GInThrGInGInGInLeuGInArgLeuArgGInGInIleMetAspLysThrArgLeuIle 1260
 QY 268 AAAGAAAGAGATGACTGTGATGCTGCTGACTGTTCCGTAAGAGACGCTGGACATACG 327
 DB 1261 GInArgCysAspLeuGInLLeuTyrGInLeuLysGInGInIleGInAlaLeuLysAspThr 1280
 QY 338 CAGCAAGAGAGCAAGCAAGCTTATGACAGAGTGAACAAGTTTGCATATATCTGAGGAA 387
 DB 1281 LysPro---GInValGInThrArgGInValValGInGInGInIleLeuGInPheGInGInLysp 1299
 QY 388 GCCAAT-----TTGAAAAAACCAAGCT 411
 DB 1300 ProGInThrLysLysGInValGInLysLeuArgLysGInLeuSerGInGInLysLys 1319
 QY 412 TTAATCCAGTGTGACAGCTGAGGAGAGAGCTGAGAGACGAGCGAGCACTTGAAAAA 471
 DB 1320 GInValAspLeuGInGInGInArgLysArgLysGInGInLysLysLysArgLysLysLys 1339
 QY 472 GAACCTGCATCTCAGCAAGAGAAAGGCCATTGGAAGAACATG----- 516
 DB 1340 GInLeuAlaGInGInArgLysGInLysGInValValArgGInGInValGInLysLysp 1359
 QY 517 -----ATGAAAAAGGAATATACG-----AAA 537
 DB 1360 GInProAspLeuArgAlaGInValThrAlaPheThrAsnSerIleAspAlaGInLeuArg 1379
 QY 538 GAAAGGAGTACATGAGATCAAGAGATGTTGATCTGCTCAGAAATATGCCCACTGGAG 597
 DB 1380 GInIleAspLysLeuIleValGInLeuArgArgLeuGInHisArgArgAlaGInLeuGIn 1399
 QY 598 GCCCAGGTGAGAAAGGTTACAAAGGAA-----AAGATTTCAGCTATTATCAACTG 648
 DB 1400 ArgGInLeuGInGInLeuGInArgLysGInAlaArgAlaArgAlaGInLeuGInVal 1419
 QY 649 GAGGAATTCAAAGCCAGCTGCTTCGGGAATGATGTCACAAAGCTGTGGAGAA 708
 DB 1420 GInArgLeuGInGInArgLeuAlaLeuGInGInGInGInGInLysThr---GInGIn 1438
 QY 709 ATGCCCTATCAGCTAATATAAACCAACATGAGAGAGAT---GAGCGAGAAAGAGAGAC 765
 DB 1439 LysValIThrIsthrGInLysValValLeuGInAspProGInGInThrArgLysHis 1458
 QY 766 AGAGAGTTCAGAGCAAAACTAACAGGAT-----CTGAATTAAGATACAGGAA 816
 DB 1459 AlaLeuLeuArgAlaGInGInGInGInGInGInGInGInGInGInGInGInGInGIn 1478
 QY 817 ATAGAGAAATTCAGATGAGAACTGATGAGAAACAAACAACTTGAACAGAGAGAGAG 876
 DB 1479 LeuGInProLeuArgArgLysLeu-----AlaAlaLeuGInLysAlaGInLys 1494

DB 1393 LysLysLeuAlaGlnArgLeuGlnAspAlaGluGlnHisValGluAlaValAsnAlaLys 1412
 QY 1132 CTGCTGACGCTCTCGAAGCAACACACCTCTCTCTGAGAGGACGCTGTGGAA 1191
 DB 1413 CysAlSerLeuGlnLysThrLysGlnArgLeuGlnAsnGluValGluAspLeuMetVal 1432
 QY 1192 GAGGTGACCGC 1203
 DB 1433 AspValGluArg 1436
 RESULT 14
 MYHA_HUMAN
 ID MYHA_HUMAN STANDARD: PRT: 1976 AA.
 AC P35580;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, last sequence update)
 DT 16-OCT-2001 (Rel. 40, last annotation update)
 DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain,
 type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).
 GN MYH10.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96025307; PubMed=7499478;
 RA Phillips C.L., Yamakawa K., Adelstein R.S.;
 RT "Cloning of the cDNA encoding human nonmuscle myosin heavy chain-B and
 analysis of human tissues with isoform-specific antibodies.";
 RL J. Muscle Res. Cell Motil. 16:379-389(1995).
 RN [2]
 RP SEQUENCE OF 63-722 FROM N.A.
 RX MEDLINE=91316803; PubMed=1860190;
 RA Simons M., Wang M., McBride O.W., Kawamoto S., Yamakawa K.,
 RA Gdula D., Adelstein R.S., Weir L.;
 RT "Human nonmuscle myosin heavy chains are encoded by two genes located
 on different chromosomes.";
 RL Circ. Res. 69:530-539(1991).
 CC -1- FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS,
 CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND
 CAPING.
 CC -1- SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY
 CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2
 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
 CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 DB EMBL: M69181; AAA99177.1; -
 DB PIR: B61231; B61231.
 DB HSSP: P10587; LBR2.
 DB Genem: HGNC:7568; MYH10.
 DB MIM: 160776; -
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR002928; Myosin_tail.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head; 1.
 DR Pfam: PF00612; IQ; 1.
 DR Pfam: PF01576; myosin_tail; 1.
 DR Pfam: PF02736; myosin_N; 1.

DR PRINTS: PR00193; MYOSINHEAVY.
 DR Prodom: PD000355; myosin_head; 1.
 DR SMART: SM00015; IQ; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ; 1.
 KW Myosin; ATP-binding; Calmodulin-binding; Actin-binding;
 FT Coiled coil; Alkylation; Multigene family;
 FT DOMAIN 1 785 MYOSIN HEAD-LIKE.
 FT DOMAIN 786 815 IQ.
 FT DOMAIN 845 1976 COILED COIL (POTENTIAL).
 FT NP_BIND 178 185 ATP (POTENTIAL).
 FT MOD_RES 701 701 ALKYLATION (SH-1) (POTENTIAL).
 FT MOD_RES 711 711 ALKYLATION (SH-2) (POTENTIAL).
 SQ SEQUENCE 1976 AA; 228938 MW; B2BB87FEF35EA124F CRC64;

Alignment Scores:

Pred. No.:	2, 65e-06	Length:	1976
Score:	240.50	Matches:	115
Percent Similarity:	46.018	Conservative:	104
Best Local Similarity:	24.16%	Mismatches:	148
Query Match:	8.80%	Indels:	109
DB:	1	Gaps:	23

US-09-502-945-1 (1-1552) x MYHA_HUMAN (1-1976)

QY 7 GATGATCCGAGAGCTAAACTTACTTATGAGGAAAGTGTAATTGAGAAATCCCAA 66
 DB 1038 Aspleuclnclunrgrleuylslysglunlgrthrglncglunleuglulysalals 1057
 QY 67 TTGAAGTT-----TTGAGACGACTTACTGATATTCAGAGAACT 108
 DB 1058 ArgLysLeuAspLysGlnThrThrAspLeuGlnAspLlnleAlaGlnleuGlnAlaGln 1077
 QY 109 TGTGAAGATCTTAAGACCACTTAAGCAATAAATTTCTTGCGCTCT----- 159
 DB 1078 llaAspLlnleuylsleuGlnleuAlaLysGlnGlnleuGlnleuGlnlAlaLeuAla 1097
 QY 160 -----AATCTGTAACCGTGTGGTCTTTGTTGAATGCTCAGCAT 207
 DB 1098 ArgGlyAspAspLlnThrThrLeuHisLysAsnAsnAlaLeuLysVal--ValArgGluLeu 1116
 QY 208 GAAGCTGTCTCTCCCAAAACCCTACTAATGTCATATGACCAATCGAAGA----- 261
 DB 1117 GlnAlaGlnlAlaGlnleuGlnGlnlAspPheGlnSerGlnLysAlaSerArgAsnLys 1136
 QY 262 CTGCTTAAGAAAGAGATGATGATGCTGCACTGATTTCCGTAAGACACTTGCA 321
 DB 1137 AlaGlnLysGlnLysArgAspLeuSerGlnGlnleuAlaLeuLysThrGlnleuGln 1156
 QY 322 GATACGACGCAAGAGACAGCAAGTCTTATGAA-----CAGGTGAACACAGTTTGCAATA 378
 DB 1157 AspThrLeuAspThrThrAlaAlaGlnGlnleuAlaArgThrLysArgGlnGlnleuAl 1176
 QY 379 TCTGAG-----GAGCCAAATTTGAAAACCAAG-----GCTTAATTCAG-- 420
 DB 1177 AlaGlnleuLysLysAlaLeuGlnGlnlGlnLysnHisGlnAlaGlnlIleGlnAsp 1196
 QY 421 -----TGTACCACTTGAAGAGAGAGCTGAGAGCGAGCG 456
 DB 1197 MetArgGlnArgHisAlaThrAlaLeuGlnleuLysSerGlnleuGlnleuGlnleuGlnleu 1215
 QY 457 GAGCGACTTGAAAAGAACTTGATCTCAGCAAGAAAAGGCCATTTGAGAAACAGT 516
 DB 1216 LysArgPheLysAlaAsnleu-----GlnLysAsnLysGlnleuGlnleuThrAsp--- 1232
 QY 517 ATCAAAAAGCAATAAAGCAAGAAAGAGGAGTACATGGGATCAAGATGTTGATCTGTCT 576
 DB 1233 ---AsnLysGlnleuAlaLysGln-----ValLysValleuGln 1244
 QY 577 CAGAAATATGCGCAACTGGAGCCAGGTGAGAAAGATTACAAAGGAAAGATTACAGCT 636
 DB 1245 GlnValLysAlaGlnSerGlnHisLys-----ArgLysLysLeuAspAla 1259

OY	637	ATTATATCAACGTGGAGGAAATAATCAAAGCACCTGGCTTCCGGGAA-----ATGGAT	687
Db	1260	-----GlnValGIngluLeuHisAlaLysValSerGluLysAspArgLeuArgValGlu	1277
OY	688	GTCACAAGAAGTGCTGGAGAAATGCCGTATACACTGAATAAACCMACTAGAGAAGAT	747
Db	1278	LeuAlaGluLysAlaSerLysLeuGlnAsnGluLeuAspAsnValSerThrLeuLeuGlu	1297
OY	748	GAGCAGCAAAAAGCAGCACACAGAGTTCTCAGCGCAAAAAACTACAGGATCTTGAATTA	807
Db	1298	GluAlaGluLysLysGlyIleLysPhe--AlaLysAspAlaAlaSerLeuGlnSerGln	1310
OY	808	GATGAGAAATATGAGAAATG-----	828
Db	1317	LeuGlnAspThrGIngluLeuLeuGlnGluIuThrArgGlnLysLeuAsnLeuSerSer	1336
OY	829	AGAAATTA---GAACATGATGTGAACCAACAACACTTGGACAGAGCAGCAAGACACCC	885
Db	1337	ArgIleArgGlnLeuGlnGluGluLysAsnSerLeuGlnGlnGlnGlnGluGluGlu	1355
OY	886	CTGGCCACGA-----	894
Db	1357	GluAlaArgLysAsnLeuGlnLysGlnValLeuAlaLeuGlnSerGlnLeuAlaAspThr	1376
OY	895	-----GAGGAGTGCCTGAGACTATACAGAACTCTGGCGCATCTACACACCA	942
Db	1377	LysLysLysValAspAspAspLeuGlyThrIleGlnSerLeuGlnGluAlaLeuLafyrsp	1396
OY	943	CTGCACCTCACACAGATCTGAATAATGAGCCAACCTGATGCACGAAAAAAGATATCATATGAT	1002
Db	1397	LeuLeuLysAspAlaGluAlaGluLeuSerGlnArgLeuGlnGluLysAlaLeuAlafyrsp	1416
OY	1003	AAATTTGGAAAAATTACAGAGAGA---AATGAAGAATTGGAGAACAGTGTCCAGCAT	1059
Db	1417	LysLeuGlnLysThrLysAsnArgLeuGlnGlnGluLeuAspAspLeuThrVal-----	1434
OY	1060	GGGAGATCACTAGACAGATGAACCAAGGCTTAAGCAGCTGGATATACACAGCCAGGCC	1119
Db	1435	-----AspLeuAsp-----HisGlnArgGln	1441
OY	1120	ACACCCACAGCTGGTGCACGCTTCACAGACAGACACGAGTCTCTCGAGAGGCG	1179
Db	1442	ValAlaSerAsnLeuGlnLysLysGlnLysLysPheAspGlnLeuLeuAlaGluLys	1461
OY	1180	AGCCTGTG-----GAGAGGTGACCGCGCTGCGGACCCAG	1215
Db	1462	SerLeuSerAlaArgTyraIaGluGlnLysArgAlaGluValGlu	1477
<hr/>			
RESULT 15			
MISP_BRUMA			
ID	MISP_BRUMA	STANDARD;	PRT; 880 AA.
AC	Q01202; P90711;		
DT	01-FEB-1996 (Rel. 33, Created)		
DT	15-JUL-1998 (Rel. 36, Last sequence update)		
DT	15-JUL-1998 (Rel. 36, Last annotation update)		
DE	Paramyosin.		
OS	Brugia malayi ¹ .		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;		
OC	Nematoidea; Brugia.		
CC	NCBI_TaxID=6279;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Lamy S., Luganda P., Nicolas L.;		
RL	Submitted (NOV-1996) to the EMBL/Genbank/DBJ databases.		
RX	[2]		
RE	SEQUENCE OF 180-263 AND 684-880 FROM N.A.		
TX	MEDLINE=92131066; PubMed=17515173;		
FT	"Li B., Chandrasekhar R., Alvarez R.M., Lifits F., Weil G.-J.;		
RT	"Identification of paramyosin as a potential protective antigen		
RL	against Brugia malayi infection in jirds."		
RL	Mol. Biochem. Parasitol. 49:315-324(1991).		
CC	-I- FUNCTION: PARAMYOSIN IS A MAJOR STRUCTURAL COMPONENT OF		
	MANY THICK FILAMENTS ISOLATED FROM INVERTEBRATE MUSCLES.		

	-1	SUBUNIT: HOMODIMER (BY SIMILARITY).	
CC	-1	SUBCELLULAR LOCATION: Thick filaments of the myofibrils.	
CC	-1	SIMILARITY: HIGH, TO MYOSIN HEAVY CHAINS.	
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on CC use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and/or commercial entities requires a license agreement (See http://www.isdb.ch/annotation or send an email to licensese@isdb.stb.ch).		
DR	EMBL; U77590;	AACI8613.1 ; -	
DR	EMBL; M63097;	AAA27859.1 ; -	
DR	HSSP; P09372;	IDKG.	
DR	InterPro:	IPIR002928; Myosin_tail.	
DR	Pfam:	PF01576; Myosin_tail; L.	
KW	Cooled coil;	Muscle protein; Thick filament;	
KW	Myosin.		
FT	DOMAIN	1 34 NONHELICAL REGION (POTENTIAL). COILED COIL (POTENTIAL).	
FT	DNAIDN	35 859 NONHELICAL REGION (POTENTIAL).	
FT	DOMAIN	860 880 V -> I (IN REF. 2).	
FT	CONFLICT	189 189 D -> A (IN REF. 2).	
FT	CONFLICT	197 197 . L -> F (IN REF. 2).	
FT	CONFLICT	251 251 E -> Q (IN REF. 2).	
FT	CONFLICT	255 255 A -> R (IN REF. 2).	
FT	CONFLICT	696 696 FRKTRKR -> IQEETA (IN REF. 2).	
FT	CONFLICT	730 736 L -> IQ (IN REF. 2).	
FT	CONFLICT	825 825 HOLLRKKMLQRRKFSPSKNRDN -> SSVVTKKNASAKS IVLEDEQ (IN REF. 2).	
FT	CONFLICT	857 880	
SO	SEQUENCE	880 AA; 101904 MW; FT9A27E64D2DA85E CRC64;	
	Alignment Scores:		
	Score No. :	3.35e+06 Length: 880	
	Percent Similarity:	239.50 Matches: 117	
	Best Local Similarity:	41.07% Conservative: 113	
	Query Match:	8.77% Mismatches: 163	
		Indels: 167	
	DB:	Gaps: 21	
US-09-502	-945-1 (1-1552) x MYSE_BRUMA (1-880)		
OY	4 CTGATGCATCGAGAACTTAACATTACTTGAGAATAAGTTGAATTCAGCAATCC	63	
Db	: ::: ::: :: :::		
133	LousapryralagluInglInleGlunLeuglnLylysamsrlyslleasprgllu	152	
OY	64 CAATTGANGTTTTGGAGAACGACTTAGGTGAATCACAAGAACTTGTAAGATTTAA	123	
Db	::: ::: :: ::::		
153	ArgclnArty--Leuginhlsglvalllleglleuthrlalthrileaspelnleuyys	171	
OY	124 GAGCACATAAAGCTTAAGAATTCTTCGGCGCTAATACATTGTAAACCCTTGCTGGT	183	
Db	: :		
172	Lysasp---LysHIS-----LeualaaglulysalaLaLacluArphgeCluaLa	186	
OY	184 CTTTGTGTAANTGCTCATGACATGAAGCGTGTTCGCCAACACCATGACTACTGTCAT	243	
Db	::::: ::::: :: :::		
187	GlnthrValGIuleuseerAnlysvAlGLUAspLeuAnahty---HisVAlasmnspleu	205	
OY	244 ATGCAGACCATCGAAGACTGGTTAAAGAAGAAGATGACTGTAGTGTCTGCACCTAGTTCC	303	
Db	:::::::::: :::: ::: :::		
206	ALacInglnArtyglnaAglnlaIglnlaIGlnslnasndAspleudeulygsIUlleHIsasp	225	
OY	304 GTAAGGAGCACTTGGCGAAGTACCGSCGAAGAAGCAAAGTGCCTTAATACAGGTGMA	363	
Db	::: :::: :::: :::: ::: :::		
226	GlntylvAlGIleuaspaanhLeughnlhsVallLystrycLnleualaeGlnleugnu	245	
OY	364 CAATTTTGCAGAAATATGTGAGGAAGCAATTTTAAAAAACAAGCGCTTAATCCAGTGT	423	
Db	::: :::: :: ::: :: :::		
246	GIUnlahrtgrAtgytleuGIUsAprlaIGluatrgLUarSerctInleuglnAleGlnleu	265	
OY	424 GACCAAGTTGAGGAAGAGtTGGAAGGACGGCAGCACTTGAAAAAAGACTTCATCT	483	

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:32:38 ; Search time 27.6459 Seconds
(without alignments)
10793.685 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732
Sequence: 1 cttctgagatgcacgcagaaa.....aaatgacttttaagaaga 1552

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+n2p.model -DEV-xlp
-Q/cgr2_1/USPTO.spool/US09502945/r/unat_14032003_101059_19127/app-query.fasta.1.10979
-DB-pir_73 -QPM-fastlan -SUFFIX-rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE-LOCAL
-OUTFM=pic -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US09502945 @CGN_1.1.263 @unat_14032003_101059_19127 -NCPU=6 -ICPU=3
-NO_XLPRY -NO_MMAP -LARGEOUTPUT -NRC_SCORES=0 -WAIT -LONGLOG -DEV-TIMEOUT=120
-MAIN-TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR_73: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	272.5	10.0	1738	2	T14867
2	264	9.7	1922	2	T00637
3	256.5	9.4	1790	2	S67593
4	255.5	9.4	978	2	A70387
5	255.5	9.4	1959	1	A33977
6	254	9.3	2017	1	A36014
7	254	9.3	2057	2	S61477
8	253.5	9.3	1931	2	A59234
9	253.5	9.3	3225	2	I52300
10	253.5	9.3	3259	1	A56539
11	252	9.2	4574	2	G02520
12	249.5	9.1	2007	1	B43402
13	248.5	9.1	4687	1	A39638
14	246.5	9.0	1992	2	A47297

15	245	9.0	2245	2	T18278	myosin heavy chain
16	244	8.9	764	2	I51302	myosin heavy chain
17	244	8.9	1940	2	A29320	myosin heavy chain
18	244	8.9	4684	2	A59404	plectin [imported]
19	242	8.9	2663	1	S28261	centromere protein
20	241.5	8.8	936	2	S39083	myosin heavy chain
21	241	8.8	1927	2	A59236	embryonic muscle m
22	241	8.8	1940	2	A59287	myosin heavy chain
23	240.5	8.8	1938	1	JX0178	myosin heavy chain
24	240.5	8.8	1976	2	A59252	myosin heavy chain
25	239.5	8.8	746	2	T47237	myosin II heavy ch
26	239.5	8.8	555	2	S24348	myosin heavy chain
27	239.5	8.8	2442	2	T08621	centrosome associa
28	239	8.7	879	2	A48575	paramyosin - nemat
29	239	8.7	1961	1	A61231	myosin heavy chain
30	238.5	8.7	1760	2	T17272	hypothetical prote
31	238	8.7	848	2	A44972	paramyosin - nemat
32	237	8.7	1937	2	I38055	myosin heavy chain
33	236.5	8.7	1934	2	I48153	myosin heavy chain
34	236	8.6	1690	2	T13030	microtubule bindin
35	236	8.6	1935	1	S06006	myosin beta heavy
36	236	8.6	1940	1	A24522	myosin heavy chain
37	235	8.6	1130	2	T34081	hypothetical prote
38	235	8.6	1972	1	A41604	myosin heavy chain
39	234	8.6	741	2	S39082	myosin heavy chain
40	234	8.6	1549	1	A40691	trichopolyalin - she
41	234	8.6	1935	2	A59286	myosin heavy chain
42	233.5	8.5	631	2	JC4298	hyaluronan recepto
43	233.5	8.5	924	2	S06117	myosin heavy chain
44	233.5	8.5	1133	2	T22976	hypothetical prote
45	233.5	8.5	1818	1	S73852	hypothetical prote

ALIGNMENTS

RESULT 1
T14867
Interaptin - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000
C:Accession: T14867
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.
J. Cell Biol. 142, 735-750, 1998
A:Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dicts.
A:Reference number: Z18248; MUID:98365468; PMID:9700162
A:Accession: T14867
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1738 <RIV>
A:Cross-references: EMBL:AF057019; NID:93549260; PID:93549261; PIDN:AMC34582.1
C:Genetics:
A:Gene: abpd
A:Introns: 173/2; 1680/1

Alignment Scores:

Score: 272.50	Length: 1738
Percent Similarity: 42.16%	Matches: 106
Best Local Similarity: 20.78%	Conservative: 109
Query Match: 9.97%	Mismatches: 162
	Indels: 133
DB: 2	Gaps: 17

US-09-502-945-1 (1-1552) x T14867 (1-1738)

QY	13	TCCGAGAACTAAAGCTTATGAGCAAAAGTGTGAATTCGAAATTCGAAG	72
		::: ::: ::: ::: ::: ::: :::	
DB	844	ASndiulysasnginllyAsarhnlsglnAspdlnduileuileuigulysGlnleu	863
QY	73	TTTTTGGAAGCAAGCACTAGCTGATATCAGAGAACTGTGAAGTCTTAAGCAACTA	132
		::: ::: ::: ::: ::: :::	
DB	864	GlnleuIngInglngulYrAspGlnLeuAsnGlnlthrasnGlnserlIleGluAsnGlnLeu	883

```

QY 133 AAGCATAAAGATTCTCTGCTGCTAACTTGTACCGCTGTGCTGCTTTG 192
DB      :      :      :      :      :      :      :      :      :
DB 884 AaNGlInGlnAspLeuInLysGlnAsn-----Leu 894
QY 193 AATATGTCACAGATGACCTGTTCTTCCAAACCCATACCTAATGTCATATGACAGCC 252
DB      :      :      :      :      :      :      :      :      :
DB 895 AaNGlULysGlnGlnLeuLeuLysLeuGlnAsnGlnLeuAsnGlnGlnLeuLys 914
QY 253 ATCGAA-----AGACTGTGTAA 270
DB      :      :      :      :      :      :      :      :      :
DB 915 lIeGlnPheAspGlnGlnIupPheSerLysGlnAsnSerLLeAsnllIeGlnLeuVala 934
QY 271 GAAAGAGATGACTTGTATGTCGACTGTTCCGTAAAGACAGCTTGGCAGATACCGAG 330
DB      :      :      :      :      :      :      :      :      :
DB 935 GlnLysAsnGln-----LysLeuLlIeGlnLeuGln 944
QY 331 CAAGAGAAAGCAAGTCTTATGACAGGTAACAAAGTTTGGCAATATCTGAGGAAGC 390
DB      :      :      :      :      :      :      :      :      :
DB 945 Gln-----AspLysAsnGlnLeuLysGlnGlnAsnArgSerAsnAspGlnLys 960
QY 391 AAT-----TTGAAAAAACCAAGGCTTTAATCCAGTGTACAGCTTGAGAGAG 438
DB      :      :      :      :      :      :      :      :      :
DB 961 AspGlnAsnAspLeuLlIeGlnLysGlnAsn-----GlnLeuLysSerLlIeGlnAsn 977
QY 439 GAGCTGAGAGCGAGCGGAGCGACTTGAAAAAGAACTTGCATCTCAGCAA---GAGAAA 495
DB      :      :      :      :      :      :      :      :      :
DB 978 GlnLeuAsnGlnLeuLlIeGlnLysAsnGlnSerAspHlLysGlnGlnGlnLeuLysGln 997
QY 496 AGGCCCATTTGACAAAGACATGATGAAAAAGAA----- 528
DB      :      :      :      :      :      :      :      :      :
DB 998 GlnSerLlIeGlnAsnAspLeuLlIeGlnLysGlnAsnGlnlIeGlnGlnLeuGlnSerGln 1017
QY 529 AAACAGCAAAAGAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 588
DB      :      :      :      :      :      :      :      :      :
DB 1018 LeuAsnGlnGlnArgGlnGlnGlnSerAsnGln-----LeuSerGlnLysAspGln 1034
QY 589 CAAGTGGAGCCCGAGGTAAG----- 612
DB      :      :      :      :      :      :      :      :      :
DB 1035 GlnLeuAsnGlnLeuLlIeGlnLysAsnGlnPheAspGlnLysGlnGlnGlnLeuLysGln 1054
QY 613 -----GTTACAAAGCAAAAGATTTACAGTATTAATCACTGAGGAAATTCAAACCCAG 666
DB      :      :      :      :      :      :      :      :      :
DB 1055 GlnSerLlIeGlnAsnAspLeuPheGlnLysGlnAsnGlnlIeGlnGlnLeuGlnSerGln 1074
QY 667 CTGGCTTCTGGGAAATGATGTCACAAAGTGTGTGAGAAATGCGCTATCAGCTGAAT 726
DB      :      :      :      :      :      :      :      :      :
DB 1075 LeuAsnGlnGlnArgGlnGlnGlnSerAsnGlnLeuSerGlnLysAspGlnGlnLeuAsn 1094
QY 727 AAACCAACATGAGAGATGAGCGACAAAGAGACACAGAGTTCAGAGCAAAACT 786
DB      :      :      :      :      :      :      :      :      :
DB 1095 GlnLeu---lIeGlnLysAsnGlnSerAspGlnLysGlnGlnGlnLeuLysGlnGlnSer 1113
QY 787 ---AACAGGATCTGTAATTTAAGATCAGGAATGAGAAATTTGATAGAACGCGAT 843
DB      :      :      :      :      :      :      :      :      :
DB 1114 lIeGlnAsnAspLeuLlIeGlnLysGlnAsnGlnlIeGlnGlnLeuGlnLeuAsn 1133
QY 844 GAAAGCAAAACA----- 855
DB      :      :      :      :      :      :      :      :      :
DB 1134 GlnGlnArgGlnLeuGlnSerGlnValSerLlIeAspAsnAspLysLlIeLeuGlnLeuGln 1153
QY 856 -----CACTTGGACAGAGACAGCAG 876
DB      :      :      :      :      :      :      :      :      :
DB 1154 LysGlnLeuLysGlnLysGlnSerAspLeuLysLeuAsnAspLlIeLysGlnGlnGln 1173
QY 877 AAGGAGCCCTGGCCAGAGAGAGTCC-----CTGAGACTACA 915
DB      :      :      :      :      :      :      :      :      :
DB 1174 AspLysGlnLeuGlnAspLysGlnlIeGlnPheAspGlnLeuGlnLeuThrPheAsnGln 1193
QY 916 -----GAACTGTGGCGAATCTGAGACCAACTG 945
DB      :      :      :      :      :      :      :      :      :
DB 1194 PheLysAsnAspLysAspSerGlnPheLlIeGlnLeuGlnAspAspGlnLysGlnGlnLeu 1213
QY 946 CACCTCACCAGATCTGAATAGCTCACTCACTCAAGAAAAAGGTATACATAT----- 999

```

```

DB 1214 GlnSerLlIeGlnLysAspLeuAsnGlnLeuLysGlnGlnGlnGlnLysGln 1233
QY 1000 -----GATTAATTTGGAAAGTTTCACAGAGAAAGAAATGAGAAATTTGAGGAA 1044
DB 1234 LeuSerGlnLysAspLlIeGlnLysGlnSerLlIeGlnPheGlnAsnGlnGlnLysGlnLys 1253
QY 1045 CAGTGTGTCACAGATGGAGAGTACATGAGACGATGAGCAAGGCTAAGGCACTGAT 1104
DB      :      :      :      :      :      :      :      :      :
DB 1254 GlnLeuSerGlnLysAspLlIeGlnLysGlnSerLlIeGlnGlnAsnGlnLeuAsn 1273
QY 1105 AAGCAGCCAGCCAGCCAGCCAGCAG-----CTGGTGCAGCTC 1143
DB      :      :      :      :      :      :      :      :      :
DB 1274 AspGlnAsnGlnGlnLysValLysGlnPheSerGlnLysAspLlIeLysGlnGlnSerLlIe 1293
QY 1144 CTCAGACAGCAGACAGGCTTCCCTGAGAGAGCAGACAGCTTGGGAGAGAGTGACCCG 1203
DB      :      :      :      :      :      :      :      :      :
DB 1294 GlnGlnAspLeuAsnGlnLysGlnGlnGlnGlnGlnGlnGlnLysGlnLysGlnLeuSerGln 1313
QY 1204 CTGCGGACCCAGTTACCCAGCATGCCACAA 1233
DB      :      :      :      :      :      :      :      :      :
DB 1314 LysAspGlnLysLeuGlnSerLlIeGlnGln 1323

```

RESULT 2

```

T00637
hypothetical protein H_G5541B18.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 05-Nov-1999
C:Accession: T00637
C:R:Hinds, K.; Sulterer, C.; Becker, M.; Hawkins, M.
submitted to the EMBL Data Library, January 1998
A:Description: The sequence of H. sapiens BAC clone G5541B18.
A:Reference number: 214196
A:Accession: T00637
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1922 <HIN>
A:Cross-References: EMBL:AC004013; NID:g2781380; PIDN:AAB96867.1; PID:g2781381
A:Genetics:
A:Map position: 7q21
A:Note: Intron positions not resolved (incomplete sequence)
A:Note: H_G5541B18.1

```

```

Alignment Scores:
Pred. No.: 3,07e-09 Length: 1922
Score: 264.00 Matches: 121
Percent Similarity: 40.57% Conservative: 107
Best Local Similarity: 21.53% Mismatches: 162
Query Match: 9.66% Indels: 172
DB: 2 Gaps: 22

```

US-09-502-945-1 (1-1552) x T00637 (1-1922)

```

QY 1 CTTCGAGATGATCCGAGAGCTAAACTTACTTATGAGAAAGTGTGAA---ATTGAG 57
DB      :      :      :      :      :      :      :      :      :
DB 51 MetLeuAsnllIeSerSerArgLeuGlnAlaValaGlnLysLeuLeuGlnAlaIleSer 70
QY 58 GAA-----TCCCAATTAAGTTTGTGAGAACGAC 87
DB      :      :      :      :      :      :      :      :      :
DB 71 GlnThrSerSerGlnLeuGlnlIeAlaLysValThrGlnThrGlnLeuMetArgGlnSer 90
QY 88 TTAGCTGAATATCAGAGAACTTGTGAGATCTTAAAGACACATAAGCATTAAGATTT 147
DB      :      :      :      :      :      :      :      :      :
DB 91 PheArgGlnLysGlnGlnLarThrGlnSerLeuLysGlnGlnGlnGlnLysArgGlnArg 110
QY 148 CTTCGCGCTGCTAATACCTTGTAAACCGTGTGGTCTTTGTTGAAATGTCTCAGCAT 207
DB      :      :      :      :      :      :      :      :      :
DB 111 LeuHlIeGlnGlnSerArgLarGln-----GlnLeuAlaValaGlnLeuSerLysAla 128
QY 208 GAAGCGTCTTCCCAAAACCCATACTATATGTTATATGACACCATGCAAGAGCTGGTT 267
DB      :      :      :      :      :      :      :      :      :
DB 129 GlnGlnValIleAspLysPylTyralaAspGlnLysThrLeu-----PheGlnArgGlnIle 146

```


QY	64	CATTGAACTTTTGTAGGACGACTTACGTGAATATCGAGAACTTGGAACTGT---	120
Db	963	LysLeuLysSerLeuAlaSnAsnLysrLysAspMetGlnAlaGluAsnLysLeuLeu	982
QY	121	-----AAGAGCAACTAAGCATTAAGAAATTTCTTGGCTGCATTAATCTGTAAACGT	174
Db	983	LysAlaValaIGLUGLuserLysAsnGlsuserSerIleGlnLeuSerAsnLeuLnsLys	1002
QY	175	GTTGGGGCTTTGTTTGAATGTGCTCAGCATGAAGCTGTTCTTCCCAACCCTACT	234
Db	1003	IleAspSer-----MetSerGlnLysGlu	1011
QY	235	AATGTTTCATATCAG-----ACCATCGAAAGCTGGTTAAAGAAAGA	276
Db	1012	AsnPhnGlnIleGlnArgLysSerIleGlnLysAsnIleGlnLysLysThrIle	1031
QY	277	GATGACTGTATGTCTGCACCTACTTCCGTAAAGACAGCTTGGCAGATACGCAAGAA	336
Db	1032	SerAspLeuGlnLglnThrLysGlnGlnLleIleSerLysSerAspSerSerLysAspGlu	1051
QY	337	GAAAGCAAGCTTATGAACAGCTGAACAAAGTTTGGCAAAATATCTGAGAGAACCAATT	396
Db	1052	TyrGlnuserGlnIleSerLeuLeuLysGlnLysLeuGlnThrAlaThrAlaAsnAsp	1071
QY	397	AAAAAACCAAGCCTTAAATCCAGTGTGCACAGTTGAGGAAGAGCTGAGAGCAG---	453
Db	1072	GluAsnValaSnLysIleSerGlnLeuThrLysThrArgGlnGlnLeuGlnAlaGlnLeu	1091
QY	454	-----CGGAGGAGCAGCT	465
Db	1092	AlaAlaTyrLysAsnLeuLysAsnGlnLeuGlnThrLysLeuGlnThrSerLysLysAla	1111
QY	466	AAAAAAGAACTTCATCTTCACAGAGAAAGAGGCGCATGGAAGAGACATGATGAAGAA	525
Db	1112	LeuLysGlnValaLysGlnAsnGlnGlnLysLeuLysGlnLysLysIleGlnLeuGlnLys	1131
QY	526	GAAATTAAGCAAGAAAGGAGATGATGGATCAAGATGTGTATCTGTCTGACAATAT	585
Db	1132	GlnAlaThrGlnThrLysGlnGlnLeuAsnSer-----	1142
QY	586	GCCCAACTGGAGCCGAGTGGAAAAAGTTACAAAGAAAGAAATTTTCAGCTTAATCAA	645
Db	1143	-----LeuArgAlaAsnLeuGlnuserLysGlnLysGlnLysGlnAspLeuAlaGln	1160
QY	646	CTGGAGGAATTCAAAGCCAGCTGCTTCGCGAAATGGATGTACAAAGCTGTGTGA	705
Db	1161	LeuLysLysTyrGlnGlnIleAlaAsnLysGlnArgGlnTyrAsnGlnLysLeuSer	1180
QY	706	GAAATGCGCTATCAGCTGAAATAAACCAACATGGAG-----AAG	744
Db	1181	GlnLeuAsnAspGlnIleThrSerThrGlnGlnGlnLysAsnGlnuserLleLysLysLysAsn	1200
QY	745	GATGAGGCAAGAAAGAGACACAGAGATTCAGAGCAAAAACCTAACAGGATCTGAAT	804
Db	1201	AspGlnLeuGlnGlnGlnValaLysAlaMetLysSerThrSerGlnGlnSerAsnLeu	1220
QY	805	AAAGATCAGGAATAGCAAAATTCAGAAATAGAACTGATGAAGAAACAAACACTTGGAA	864
Db	1221	LysLysSerGlnIleAspAlaLeuAsnLeuGlnIleLysGlnLeuLysLys-----Lys	1238
QY	865	CAGGAGCAGACAGAGGAGCCCTG-----GCCAGAGAGAGAGTCCCTG	906
Db	1239	AsnGlnThrAsnGlnAlaSerLeuLeuGlnuserLleLysSerValaGlnuserGlnThrVal	1258
QY	907	AGACTTAACGAATCTGCGGAGATCTGACCAACACTGACCTCAGACAGATCTGAATA	966
Db	1259	LysIleLysGlnLeu-----GlnAspGlnCysAsnPhnLysGlnLysGlnVal	1274
QY	967	GCTCAACTCAGTCAGAAAAAGGTATATCATATGATTAATTTGGGAAG-----TTA	1017
Db	1275	SerGlnLeuGlnAspLysLeuLysAlaSerGlnAsnLysSerLysAsnSerLysTyrLeuGlnLeu	1294

```

Db 287 AsplyslysleuthrGluLeuValArgLysasnLysleuthrLysGluLeuAlaVal 306
QY 274 ---AGAGATGACTG-----ATGCTGACAGACTTCGCTAAGAGC 312
Db 307 LeuLysaspGluLeuSerPheAlaGlnGluLeuLysnArgIleGluAlaGlnLysGlu 326
QY 313 AGCTTGGCAGATACGACAAAGAGAA-----GCCAGTCTTATGAA 354
Db 327 LysPheLysGlnGluLysGluArgGlnLysGluLeuGlnHisArgLysLysLeuGln 346
QY 335 CAGGTGAACAGATTG---CAATATCTGAGAGACCAATTTGAAAAACCAAGCT 411
Db 347 GlnLysGlnGluLeuLysGlnLeuSerGlnLeuSerSerLysLysGlnLysGln 366
QY 412 TTAATCCAGTGTGACAGTTCGAGAGAGAGCTGAGAGCGGCGAGCGCTGAA--- 468
Db 367 Arg---GlnTyrGlnGlnAlaLysGlnGlnPheGlnLysPheSerGlnAlaGlnLys 385
QY 469 ---AAGAACTGTCATCTCAGCAAGAGAAAGGCCATTCGAAAGACATGATGAAAG 525
Db 386 GlnLysLysLeuValAlaGlnThrGlnLysLys---LeuGlnLys-----LleLys 401
QY 526 GAAATACGAAAGAAAGAGAGTACATGGATCAAGATG-----TTGATC 570
Db 402 GlnLeuPheSerGlnGlnLysGlnLysThrSerLeuLysMetLysGlnArgLeuValGln 421
QY 571 TTGCTCGAGATATGCCCCAAGCTG-----GAGGCCAGTGGAAAGGTTACA--- 618
Db 422 LeuGlnArgLysLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 441
QY 619 -----AAGAAAGATG-----TCAGCTATTAATCAACGAGAAATTCAA 660
Db 442 LysTyrLysGlnLysLysLysValHisGlnLysValLeuLysnGlnLysGlnLysGln 461
QY 661 AGCAGCTGCTTCGCGGAATG-----684
Db 462 ArgGlnLysGlnLysGlnLysGlnLysHisAlaHisMetValAlaSerTyrLeuSer 481
QY 685 -----GATGCAACAAGGTGTGAGAAATGCGCTATGAC-----CTGAATAAA 729
Db 482 ProGlnLysPheLysProValCysGlnGlyLe---TyrArgGlnLysAlaLeuGln 500
QY 730 ACCAATGAGAGAGATGAGAGCAGAAAGAGCAGACAGATTCAGAGCAAAATTAAC 789
Db 501 ValAspLysGlnLysLysLysLysLysHisAlaLysGlnLysGlnLysGlnLysGln 520
QY 790 AGGAGTTTGAATTA-----GATCAGGAA 816
Db 521 ArgGlnLysPheLysPheLysLysLysLysLysLysLysLysLysLysLysLysLys 540
QY 817 ATGAGAAATTTGAGATAGAACTGATGAAAGCAACACTTGAAACAGAGCAGCAG 876
Db 541 MetGlnLysLeuArgAsnGlnValGlnGlnLysLysLysLysLysLysLysLysLys 559
QY 877 AAGCAGACCTTGCGCAGAGAGAGTGCCTGAGATACAGAACTGCGGCAATTCAG 936
Db 560 LysGlnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 577
QY 937 CACCAACTGCACCTCACCAGATGTAATAGCTCACTCACTCAAGAAAGGATATACA 996
Db 578 HisLysLeuLysnLysTyrArgLysAlaLeuGlnLysPheGlnLysGlnLysGln 597
QY 997 TATGATAAATTTGGAAATGATAGAGAAAGAAATGAAATTTGGAGCAAGCTGCTCAG 1056
Db 598 GlnAlaLysLeuHisLysAlaGlnThrGlnLeuGlnLysLysLysLysLysLysLys 617
QY 1057 CATGGAGAGTACATGAGAGATGAACAA-----AGCTAAGGCGCTGATTAACAC 1110
Db 618 LysSerArgLeuValLysGlnPheLysGlnLysLysLysLysLysLysLysLysLys 637
QY 1111 AGCAGAGCCACAGCCAGCAG-----CTGGTCAGAGCTCTCAGC 1149
Db 638 GlnGlnSerLeuLysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLys 657

```

```

QY 1150 AAGCAGAACCGCTTCCTGCTGAGAGAGCGCTGTCGGAAGAGTGGACCGGCTCGG 1209
Db 658 LysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 677
QY 1210 ACCCAGTTACCAGCATGCGCAATCT 1236
Db 678 GlnLysLeuSerAlaLeuAsnGlnLys 686

```

RESULT 5

```

A33977
myosin heavy chain, nonmuscle - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C>Date: 31-Dec-1993 #sequence, revision 31-Dec-1993 #text, change 19-Apr-2002
C:Accession: A33977, S06116, A43422
R:Shone, R.V.; Conli, M.A.; Kawamoto, S.; Preston, Y.A.; Brill, D.A.; Adelstein, R.S.
Proc. Natl. Acad. Sci. U.S.A. 86, 7726-7730, 1989
A>Title: Cloning of the cDNA encoding the myosin heavy chain of a vertebrate cellular
A:Reference number: A33977; MUID:90046668; PMID:2813355
A:Accession: A33977
A:Molecule type: mRNA
A:Residues: 1-1959 <SHO>
A:Cross-references: GB:M26510; NID:g212382; PIDN:AAA48974.1; PID:g212383
R:Katsuragawa, Y.; Yanagisawa, M.; Inoue, A.; Masaki, T.
Eur. J. Biochem. 184, 611-616, 1989
A>Title: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed
S.
A:Reference number: S06116; MUID:90032648; PMID:2806244
A:Accession: S06116
A>Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 716-1008 <KAT>
A:Cross-references: GB:X17589
A>Note: this translation is not annotated in GenBank entry G0MCFMNA, release 114
R:Hodge, T.P.; Cross, R.; Kendrick-Jones, J.
J. Cell Biol. 118, 1085-1095, 1992
A>Title: Role of the COOH-terminal nonhelical tailpiece in the assembly of a vertebra
A:Reference number: A43422; MUID:92381096; PMID:1512291
A:Accession: A43422
A>Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1900-1959 <HOD>
A:Experimental source: brush border
A>Note: sequence extracted from NCBI backbone (NCBIP:111947)
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; ATP; coiled coil; hydrolase; methylated amino acid; nucleo
F:84-764/Domain: myosin motor domain homology <CMOT>
F:174-181/Region: nucleotide-binding motif A (P-loop)
F:552-565/Region: actin binding #status predicted
F:626-640/Region: actin binding #status predicted
F:837-1936/Domain: coiled coil #status predicted <COI>
F:837-1277/Region: S2
F:1278-1959/Region: light meromyosin
F:1937-1959/Domain: carboxyl-terminal
F:125/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:180/Binding site: ATP (Lys) #status predicted
F:694,704/Active site: Cys #status predicted

```

Alignment Scores:

```

Pred. No.: 1,05e-08 Length: 1959
Score: 255.50 Matches: 117
Percent Similarity: 43.93% Conservative: 93
Best Local Similarity: 24.48% Mismatches: 159
Query Match: 9.35% Indels: 109
DB: 1 Gaps: 20

```

US-09-502-945-1 (1-1552) x A33977 (1-1959)

```

QY 1 CTCTGATGATCCGGAAGCTTAATCTTATGAGAAAGTGTGAATTTGAGAA 60
Db 1029 TlthrAspLeuGlnLysArgLysArgGlnLysGlnArgGlnLysGlnLys 1048

```

[illegible]

```

Db 1388 LysLysLeuGlnLysAspLeuLysLeuSerLeuThrGlnArgTyrGluGluLysLysLeuAla 1407
      :::::::::::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 997 TATGATTAATTTGGGAAGTTCACAGAGAAGA---AATGAGAATTTGGAGAACAGTGTCTC 1053
      ||||||| ||| :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 1408 TyrAspLysLeuGlnLysThrLysThrArgLeuGlnGlnGlnLysAspLysLysLeuAlaVal 1427
      :::::::::::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 1054 CAGCATGGGAGAGTACATGAGACGATGAAGCAAGAAAGTAAAGCAGCAGTATAGCAGACG 1113
      :::::::::::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 1428 -----AspLeuAsp---HisLeu 1432
      :::::::::::::::::::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::
QY 1114 CAGGCCACAGCCACAGCTGTGACGCTCTACAGCAAGACAGACACCTTCTCTGGAG 1173
      :::: ||| ||| :::: :::: :::: :::: :::: :::: :::: :::: ::::
Db 1433 ArgGlnThrValSerAsnLeuGlnLysGlnLysGlnLysPheAspGlnLeuLeuAlaGlu 1452
      :::::::::::::::::::: GAAGGAGTGAGACCGGCTGCGGAGACCAG 1215
QY 1174 AGGCAAGCGCTTCG----- 1453
      :::::::::::::::::::: ||||| ||||| :::: ::::
Db 1453 GlnLysAsnLysSerAlaLysTyrAlaGlnLysAspArgAlaGluAlaGlu 1470
      :::::::::::::::::::: ||||| ||||| :::: ::::

RESULT 6
A36014
myosin heavy chain, nonmuscle - fruit fly (Drosophila melanogaster)
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Drosophila melanogaster
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
C:Accession: A36014; B36014
R:Ketchum, A.S.; Stewart, C.T.; Stewart, M.; Kiehart, D.P.
Proc. Natl. Acad. Sci. U.S.A. 87, 6316-6320, 1990
A:Title: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript.
A:Reference number: A36014; MUID:90349606; PMID:2117279
A:Accession: A36014
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-2017 <K>
A:Cross-references: GB:M35012
C:Genetics:
A:Gene: FlyBase:zip
A:Cross-references: FlyBase:FBgn0005634
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolase; methyl
F.1-2017/Product: myosin heavy chain, form I #status predicted <MA1>
F.46-2017/Product: myosin heavy chain, form II #status predicted <MA2>
F.135-815/Domain: myosin motor domain homology <MOT>
F.125-232/Region: nucleotide-binding motif A (P-loop)
F.598-621/Region: actin binding #status predicted
F.702-721/Region: actin binding #status predicted
F.888-2017/Domain: coiled coil #status predicted <COI>
F.1329-2017/Region: light meromyosin
F.176/Modified site: N6/N6, N6-trimethyllysine (Lys) #status predicted
F.231/Binding site: ATP (Lys) #status predicted
F.745/755/Active site: Cys #status predicted

Alignment Scores:
Pred. No.: 1,31e-08 Length: 2017
Score: 254.00 Matches: 110
Percent Similarity: 43.15% Conservative: 101
Best Local Similarity: 22.49% Mismatches: 174
Query Match: 9.30% Indels: 104
DB: 1 Gaps: 18

US-09-502-945-1 (1-1552) x A36014 (1-2017)

QY 4 CTTGGATGCATCCGAAACCTAAACTTATATGAGAAAGTGTGAATTTGAGGAATCC 63
      ::::| ||| ||| ::::|
Db 1143 IleAspGlnGluSerAlaThrLysAlaThrAlaGlnLysAlaGlnArgGluLeuGlnLys 1162
      :::::::::::::::::::: GCTGAATATAG 102
QY 64 CAATGAAGTTTTCAGACGACTTA----- 1163
      ||||| ::::| |||||
Db 1163 GlnLeuAlaGlnLysGlnLysAspLeuLysAlaGlnLysAlaAlaArgAlaLysAlaGlu 1182
      :::::::::::::::::::: CATTAAGAATTTCTTGCGCT 156
QY 103 AGAAGTTGTGAAGTCTTTAAAGAGCACTAAAG-----CATTAAGAATTTCTTGCGCT 156
      ::::| ||||| ||||| ::::|

```

Db 1183 LysValAlaGArgAspLeuSerGluGluLeuAlaLeuLysAsnGluLeuLeuAspSer 1202
 QY 157 GCTAATCTGTAACCGGTGGTGTCTGTTGAATGCTGCAG----- 204
 Db 1203 LeuAspThrThrAlaAlaGlnGlnGlnLeuArgSerLysArgGluGlnGluLeuAlaThr 1222
 QY 205 -----CATGAAGCTGTCTT----- 219
 Db 1223 LeuLysLysSerLeuGluGluGluThrValAsnHisGluGluValLeuAlaAspMetArg 1242
 QY 220 -----TCCAAACCCCACTAATGTTCAATTCAGACCCTCGAAAGCTGGTTAA 270
 Db 1243 HisLysHisSerGlnGluLeuAsnSerLLeaSnAspGln---LeuGluAsnLeuArgLys 1261
 QY 271 -----GAAAGAGATGACTGTGATG 288
 Db 1262 AlaLysThrValLeuGluLysAlaLysGlyThrLeuGluAlaGluAsnAlaAspLeuAla 1281
 QY 289 TCTGCACCTAGTTTCCGTAAAGAGCAGCTTGCGAGTATACGCAAGAAAGAAAGTCT 348
 Db 1282 ThrGluLeuArgSerValAsnSerSerArgGlnGluAsnAspArgArgLysGlnAla 1301
 QY 349 TATGACAGCTGGAACAAGATTTCGAATATCTGAGAGCCAAATTTGAAAAACCAAG 408
 Db 1302 GluSerGlnThrAlaGlu---LeuGlnVal---LysLeuAlaGluThrLeuAlaArg 1319
 QY 409 GCTTAATCCAG-----TGTGACCACTGAGGAAGAGCGTGCAGAGCGCAGCGAGCA 462
 Db 1320 SerLeuLeuGlnGluLysCysThrLysLeuGlnGlnGluAlaGluAsnThrAsnGln 1339
 QY 463 CTGGAAGAA-----GAACTTGATCTCAGCAAGAA 495
 Db 1340 LeuGluGluAlaGluLeuLysAlaSerAlaAlaValLysSerAlaSerAsnMetClnSer 1359
 QY 496 AGGCCCATGGAAGACATGATGAAAAAGAAATTAACGAAAGAGGAGCATACATGCGA 555
 Db 1360 GlnLeuThrGlnAlaGlnGlnLeuGlnGluGluThrArgGlnLys----- 1375
 QY 556 TCAGAAATGTTGATCTGCTCAGAAATATGCCCACTGAGCGCCAGCGTGAAGAAAGTT 615
 Db 1376 -----LeuGlyLeuSerSerLysLeuArgGlnGlnGlnGlnGlnGlnGlnGln 1392
 QY 616 ACAAGAGAAAGATTTGACGTATTATCACTGAGAAATTCAGAGCCAGCTGCTCTCT 675
 Db 1393 GlnGlu-----GlnLeuGluGluAspAspGluAlaLysArgAsn 1405
 QY 676 CGGGAATGCTGCACAAAGCTGTGTGAGAAATGCGCTATCAGCTGAATTAACCAAC 735
 Db 1406 TyrGluThrGlyLeuAlaGluValThrThrGlnMetGln---GluThrLysLysAla 1424
 QY 736 ATGAGAGAGATGAGGCGAAGAGAGACAGAGAGTTCACAGCAAGAAACCTAACAGGAT 795
 Db 1425 GluGluAspAlaAspLeuAlaLysGluLeuGlnGluGlnLysLysArgLeuAsnLysAsp 1444
 QY 796 CTGGAATTAAGATTCAGAAATAGAGAAATAGAGAAATAGAGAAATAGAGAAATAG 846
 Db 1445 IllegluAlaLeuGluValArgGlnValLysLeuLeuAlaGlnAsnAspArgLeuAspLys 1464
 QY 847 AGCAAAACAACCTTGAACAGAGCAGCAGCAGAGCAGC----- 885
 Db 1465 SerLysLysLysIlleglnSerGlnLeuGlnAspAlaThrThrIllegluAlaGlnArg 1484
 QY 886 -----CTGGCCAGAGAG 897
 Db 1485 ThrLysValLeuGluLeuGluLysLysGlnLysAsnPheAspLysIlleAlaGlnGlu 1504
 QY 898 GAGTGCCTG-----AGACTAACAGAACTGCTGGCGAATGTCAGCAACCACTCACCTC 951
 Db 1505 LysAlaThrSerGlnGlnThrAlaGlnGlnArgAspThrAlaGluValArgGluAlaArgGlu 1524
 QY 952 ACCGATCTGAAATAGCTCAACTCACTCAGCAAGAAAAAGGATACATATGATTAATGGGA 1011
 Db 1525 LysGluThrLysValLeuSerValSerArgGluLeuAspGluAlaPheAspLysIlleGlu 1544

QY 1012 AAGTTACAGAGAAATGAGAAATTTGAGAGACAGTGTCTCCAGCATGGAGATACAT 1071
 Db 1545 AspLeuGluAsnLysArgLysThrLeuGlnAsnGlnLeuAspAspLeuAlaAsnThrGln 1564
 QY 1072 GAGACATGAGCAAGAGGCTTAAGCAGCTGATTAAGCAGCAGCCAGCCAGCCAGCAG 1131
 Db 1565 GlyThrAlaAspLysAsnValHisGlnLeuGlnLysAlaLysArgAlaLeuGlnSerGln 1584
 QY 1132 CTGTGACGCTCTTCAGACAGACAGCAGCTTCTCTGGAGAGAGCAGCAGCTTCGGA 1191
 Db 1585 LeuAlaGlu---LeuLysAlaGlnAsnGlnGluLeuGlnAspAspLeuGlnLeuThrGln 1603
 QY 1192 GAGGTGACGCGCTGCGAGCAGCTTA 1218
 Db 1604 AspAla---LysLeuArgLeuGluVal 1611
 RESULT 7
 S61477
 myosin II heavy chain, non-muscle - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 19-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 02-Feb-2001
 C:Accession: S61477; S65349
 R:Manuscript, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
 J. Mol. Biol. 255, 98-109, 1996
 A:Title: Molecular organization and alternative splicing in zipper, the gene that enc
 A:Reference number: S61477; MUID:96144835; PMID:8568878
 A:Accession: S61477
 A:Molecule type: DNA
 A:Residues: 1-2057 <MAN>
 R:Manuscript, S.G.; Al-Shirawi, D.Y.; Ketchum, A.S.; Newbern, E.C.; Kiehart, D.P.
 submitted to the EMBL Data Library, September 1995
 A:Reference number: S65349
 A:Accession: S65349
 A:Molecule type: DNA
 A:Residues: 1-1908, 'NL', 1911-2057 <MAN>
 A:Cross-references: EMBL:U35816; NID:G1141789; PIDN:AB09049.1; PID:G1572481
 C:Genetics:
 A:Gene: zip
 A:Cross-references: FlyBase:FBgn0005634
 A:Interons: 10/3; 90/3; 162/3; 215/1; 255/3; 263/3; 303/3; 334/3; 1248/1; 1349/3; 1526
 C:Superfamily: myosin heavy chain; myosin motor domain homology
 C:Keywords: alternative splicing; ATP; nucleotide binding; P-loop
 F:135-855/Domain: myosin motor domain homology <MOT>
 F:225-232/Region: nucleotide-binding motif A (P-loop)
 Alignment Scores:
 Pred. No.: 1,31e-08 Length: 2057
 Score: 254.00 Matches: 110
 Percent Similarity: 43.15% Conservative: 101
 Best Local Similarity: 22.49% Mismatches: 174
 Query Match: 9.30% Gaps: 104
 DB: 2
 US-09-502-945-1 (1-1552) x S61477 (1-2057)
 QY 4 CTGATGCTATCCGAGAGCTTAAGCTTAATCTATGAGAAAGTGTGAATTTGAGAAATCC 63
 Db 1183 IleAspGlnGluSerAlaThrLysAlaThrAlaGlnLysAlaGlnArgGluLeuGlnSer 1202
 QY 64 CAATTGAAGTTTGTGAGAACGACTTA-----GCTGAATATCAG 102
 Db 1203 GlnLeuAlaGluThrLeuGlnGluAspLeuGluAlaGluLysAlaAlaArgAlaLysAlaGln 1222
 QY 103 AGAAGCTTGAGATCTTAAAGACCACTAAG-----CATAAAGATTTCTTCTGCT 156
 Db 1223 LysValAlaGArgAspLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1542
 QY 157 GCTAATCTGTAACCGGTGGTGTCTGTTGTAATGCTGCAG----- 204
 Db 1243 LeuAspThrThrAlaAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1262

```

QY 205 -----CATGAGCTGTTCT----- 219
Db 1263 LeuLysSerLeuGluGluThrValAsnHisGluLysAlaAspMetArg 1282
QY 220 -----TCCAAACCCATGATGTCATATGACACCATCGAAAGACGGTTAA 270
Db 1283 HisLysHisSerGluLeuAsnSerIleAsnAspGln---LeuGluAsnLeuArgLys 1301
QY 271 -----GAAAGAGATGACTTGATG 288
Db 1302 AlaLysThrValLeuGluLysAlaLysGlyThrLeuGluLysAlaAsnAlaAspLeuAla 1321
QY 289 TCTGCACTACTTCCGTAAGAGACAGCTGGCAGATACGACGACGACGACGAGTCT 348
Db 1322 ThrGluLeuArgSerValAsnSerSerArgGlnGluAsnAspArgArgLysGlnAla 1341
QY 349 TATGACAGAGTGAACAAAGTTTGCAGATATCGAGACCCCAATTTGAAAAACCAAG 408
Db 1342 GluSerGlnIleAlaGlu---LeuGlnVal---LysLeuAlaGluIleGluArgAlaArg 1359
QY 409 GCTTTAATCCAG-----TGTGACCACTTGAGAGAGAGCTGGAGAGCGGACGCA 462
Db 1360 SerGluLeuGlnGluLysCysThrLysLeuGlnGlnGluLysAlaAsnIleThrAsnGln 1379
QY 463 CTGAAAAA-----GAACTGCATCTCAGCAAGAGAAA 495
Db 1380 LeuGlnGluLysAlaGluLeuLysAlaSerAlaAlaValLysSerIleAsnMetGluSer 1399
QY 496 AGGCACTGAGAAAGACATGATGATAAGAAATACGAAAGAAAGAGAGATCATGGGA 555
Db 1400 GlnLeuThrGluLysAlaGlnGlnLeuLeuGlnGlnGluThrArgGlnLys----- 1415
QY 556 TCAAGAGTGTGATCTTGTCTCAGAAATATGCCCCAATCGAGCCGACGTAAGAGTT 615
Db 1416 -----LeuGlyLeuSerSerLysLeuArgGlnIleGluSerGluLysGluAlaLeu 1432
QY 616 ACAAGAGAAAGATTCAGCTATTATCACTGAGAGAAATTAACACCGCTGCTCT 675
Db 1433 GlnGlu-----GlnLeuGlnGluLysAspGlnAlaLysArgAsn 1445
QY 676 CGGAAATGATGATGTCACAAAGTGTGTGAGAAATGCGCATACGATGAATAAACCAAC 735
Db 1446 TyrGlnArgLysLeuAlaGluValThrThrGlnMetGln---GluIleLysLysLysAla 1464
QY 736 ATGAGAGAGATGAGCAGCAAGAAAGAGACACAGAGTTTCAGACCAAACTAACAGGAT 795
Db 1465 GlnGluAspIleAspLeuAlaLysGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGln 1484
QY 796 CTGGAATTAATTAATGATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAAT 846
Db 1485 IleGlnAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1504
QY 847 AGCAACACACACTTGGACAGCAGAGCAGAGCAGAGCAGCAGCAGCAGCAGCAGCAGC 885
Db 1505 SerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1524
QY 886 -----CTGGCCAGAGAG 897
Db 1525 ThrLysValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1544
QY 898 GAGTGCCTG-----AGACTAACAGAACTGCTGGGCAATCTGAGCAGCAACCTGCACT 951
Db 1545 LysAlaIleSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1564
QY 952 ACCAGATCGAATTAAGTCAACTGACGTAAGAAAGAAAGATACATATGTAATATGGCA 1011
Db 1565 LysGlnThrLysValLeuSerValSerArgGlnLeuAspGlnAlaIleAspLysIleGln 1584
QY 1012 AAGTTACAGAGAAAGAAATGAGAAATGAGAAATGAGAAATGAGAAATGAGAAAT 1071
Db 1585 AspLeuGlnLysAsnLysArgLysThrLeuGlnAsnGlnLeuAsnAspLeuAlaAsnThrGln 1604
QY 1072 GAGAGCATGAGCAAGAGCTTAAGCAGCTGATTAACACAGCCAGCCAGCCAGCAG 1131

```

```

Db 1605 GlyThrAlaAspLysAsnValHisGluLeuGlnLysAlaLysArgAlaLeuGluSerGln 1624
QY 1132 CTGGTCGAGCTCCGTCACCAAGACAGACAGCTCTCTGCGAGAGCGACAGCTGGCA 1191
Db 1625 LeuAlaGlu---LeuLysAlaGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1643
QY 1192 GAGGTGAGCCGCTGCGAGACCAAGTTA 1218
Db 1644 AspAla---LysLeuArgLeuGlnVal 1651

RESULT 8
A59234
slow myosin heavy chain 3 - quail
C:Species: Coturnix coturnix
C:Date: 19-May-2000 #sequence_rev1sion 19-May-2000 #text_change 08-Sep-2000
C:Accession: A59234
R:Nikoyts Jr., W.; Wang, G.F.; Feldman, J.L.; Miller, J.B.; Wade, R.; Nelson, L.; St
J. Biol. Chem. 271, 17047-17056, 1996
A:title: Isolation and characterization of an avian slow myosin heavy chain gene expr
A:Reference number: A59234; MVID:96291845; PMID:8663323
A:Accession: A59234
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-1931 <NIK>
A:Cross-references: GB:053862; NID:g1289513; PIDN:A659912.1; PID:g1289514
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:81-761/Domain: myosin motor domain homology <MMO>

Alignment Scores:
Pred. No.: 1,41e-08 Length: 1931
Score: 253.50 Matches: 123
Best Similarity: 39.66% Conservative: 86
Best Local Similarity: 23.34% Mismatches: 181
Query Match: 9.28% Indels: 137
DB: 2 Gaps: 20

US-09-502-945-1 (1-1552) x A59234 (1-1931)
QY 19 AGCTAAACTTACTTATGAGAAAGTGTGAATTTGAGGATCCCAATGGAAGTTTGTG 78
Db 1235 LysAlaLysAlaAsnLeuGlnLysMetCysArgSerThrGlnAspGlnMetAsnGlnHis 1254
QY 79 AGGACGACTTACCTGAATTCAGAGAACTTGTGAACAT----- 117
Db 1255 ArgAsnLysLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1274
QY 118 -----CTTAAAGACAACTAAGCATTAAGCAATTTCTTCTG--- 153
Db 1275 LeuGlnThrGluAsnSerGlnLeuSerArgGlnLeuGlnGlnGlnGlnGlnGlnGlnGln 1294
QY 153 ----- 153
Db 1295 GlnLeuMetArgGlnLysLeuThrTyrThrGlnGlnLeuGlnAspLeuLysArgGlnLeu 1314
QY 154 -----GCTGCTAATCTGTAACCGTGTGCTGTGCTGTTGTAATGCTGTCAGCAT 207
Db 1315 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1333
QY 208 GAAGCTGTTCTTCCCAACCCATTAATGTTTCATATGACGACCAATCGAAAGACTGGT 267
Db 1334 AspCysAspLeuLeuArg-----GluGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGln 1346
QY 268 AAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 321
Db 1347 GlnAlaLysAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1366
QY 322 -----GATACGACGACAAAGAGAA-----GCAAGTGCCTTAAGAA 354
Db 1367 ArgThrLysTyrGlnThrAspAlaIleGlnArgThrGlnGlnGlnGlnGlnGlnGlnGlnGln 1386
QY 355 CAGGTGAACAAAGTTTGAATATATCTGAGAGAGC----- 390

```

D	1387	LysLeuAlaGlnArgLeuGlnGluAlaGlnGluAlaValGluAlaValAlaValAlaValAlaValScys	1406
O	391	---AAATTTGAAAAAACCAAGCGCTTAATCCAG-----	420
D	1407	SerSerLeuGlnIuTrpLysThrLysSHisArgLeuGlnAsnGluIleGluIuAspLeuMetAlaAsp	1426
O	421	-----TGTGACACG	429
D	1427	ValGluArgSerAsnAlaAlaAlaAlaAlaLeuAspLysGlnArgAsnLeuAspLys	1446
O	430	TTGAGGAAGGAGCTGGAGAGCAGCGCGAGCGACTTGAAAAAGACTT---GCATCTCAG	486
D	1447	IleLeuSerGlnIuTrpLysGlnLysPheGluGlnSerGlnThrGluIleGlnValSerGln	1466
O	487	CAAGCAAAAAAGCGCCATTGAGAAAAGCATGATGAA-----AAGCAATA	531
D	1467	LysGlnAlaArgSerLeuSerThrGluLeuPheLysLeuLysAsnAlaTrpGluIuIuSer	1486
O	532	ACGAAAGAAAAGGAGTCAATGCGGATCAAAAGATTTGATCTGTCTCAGAAATTTGCCCCA	591
D	1487	LeuGlnHisLeuGlnIuThrPheLysArgGluAsnLysAsnLeuGlnGlnIuIleSerAsp	1506
O	592	CTGAGAGCCCAAGGTGGAAAAGTTTACAAGAAGAAAGATTTCACGCTAATTATCACTGGAG	651
D	1507	LeuThrGlnGlnLeuGlnIuArgLysSerGlnLys-----SerIleHisGlnLeuGln	1522
O	652	GAATATTCAAACCACGCTGCTTCGCGGAATGGATGTCCAAAGAGTGTGGAGAAATG	711
D	1523	LysValArgLysGlnLeuAspAlaGluLysLysGlnIleGlnAlaAlaLeuGlnGluAla	1542
O	712	CGTATCAGCTGAATTAACAACCAATCATGAGAGAGATGAGAGCAAGAAAGACACAGACAG	771
D	1543	GluIuAspSerLeuGlnHisGlnGlnGlnLysLysIleLeuArgAlaGlnLeuGlnIuPheAsnGln	1562
O	772	TTGAGAGCAAAAACCTAACAGCGATCTTGAAATTAAGATCAGGAATATGAGAAA-----	825
D	1563	ValLysAlaGluTrpGluArgLysLeuAlaGluLysAspGluGlnMetSerGlnIuSerLys	1582
O	826	-----TTGACAATTAAGCATGTGATGAAAGCAAAACAACCTTGAGACAGACAGACAG	876
D	1583	ArgAsnHisLeuArgVal---ValAspSerLeuGlnThrSerLeuAspAlaGluThrArg	1601
O	877	AAGGACACCCCGGCCAAGAGAGAGTCCCTGAGACTAACGAACGT-----CTG	924
D	1602	-----SerAlaGlnGlnAlaLeuAlaArgLeuLysLysLysMetGlnIuLysLeu	1617
O	925	GCGCAATCTGAGCAGCAACATG---CACCTCACCAGATGTGAATATGCTCAACTCAGTCA	981
D	1618	AsnGlnMetGlnIleGlnLeuSerHisAlaAsnAlaArgThrAlaAlaGlnAlaGlnLysGln	1637
O	982	GAAAAA-----AGTATACATATGATTAATTTGGAAAAGTTTACAGACA	1033
D	1638	ValLysAlaLeuGlnGlnIuTrpLeuLysAspThrGlnLeuGlnLeuAspAlaValArg	1657
O	1024	AGAATTAAGAATTTGGAGGACAAGCTGT-----	1050
D	1658	AlaAsnGlnAspLeuLysGlnAsnIleAlaIleValGluArgArgAsnLeuLeuGln	1677
O	1051	-----GTCCAGCATGGGAGTACATGAGACGATGAACCAAGGCTAAGCAGCTGGAT	1104
D	1678	SerGlnLeuGlnGlnIuLeuArgIleMetValGlnGlnSerGlnArgAlaArgLysLeuAla	1697
O	1105	AAGCAACCAAGGCGCACAGCCCAAGCAGCTGTGGTACAGTCTTACGCAACAGCAAC---CAG	1161
D	1698	GlnGlnGlnIuLeuThrGlnAlaSerGlnArgValGlnIleLeuLeuHisSerGlnAsnThrSer	1717
O	1162	CTTCTCTCGAGAGGACAGACCCGTGTGGAAAGAGTGAACGAGGCTGAGGACAGTATACC	1221
D	1718	LeuIleAsnGlnLysLysLysMetGlnAlaAspIleSerGlnLeuGlnThrGluValGlu	1737
O	1222	AGCATGCCACAATCTGATTGC	1242
D	1738	GluAlaIleGln---GlnCys	1743

RESULT 9
I52300
giantin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 20-Aug-1999
C:Accession: I52300
R:Sonida, M.; Mishima, Y.; Fujitwara, T.; Nishioaka, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A>Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized
A:Reference number: I52300; MUID:95100974; PMID:7802676
A:Accession: I52300
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
A:Cross-references: GB:D25542; NID:g662389; PIDN:BA05025.1; PID:g808869
C:Superfamily: giantin

Alignment Scores:	
Pred. No.:	1 4e-08
Score:	253.50
Percent Similarity:	40.33%
Best Local Similarity:	21.52%
Query Match:	9.28%
DB:	2
US-09-502-945-1 (1-1552) x 152100 (1-3225)	
	Length: 3225
	Matches: 119
	Conservative: 104
	Mismatches: 179
	Indels: 151
	Gaps: 24

US-09-502-945-1 (1-1552) x 152300 (1-3225)

Oy	1	CTTCTGATGCATCCGGAAGACGTAAACTTACTTAT-----GAG	39
Db	2170	VallelaspGluAlaLysLysTrpGluArgLysPheSerAspAlaIleGlnSerLysGlu	2189
Oy	40	GAAGAAGTGTGAATTTGGAGATCCCAATTGAACTTTTGGAGACAGCACTTACCTGAATAT	99
Db	2190	GluGluIleArgLeuLysGluAspAsnCysSerValIleuLysAspGlnLeuArgGlnMet	2209
Oy	100	CAGAGAACTGTGTGAAGTCTTAA-----GACCACTTAAGCAT-----	138
Db	2210	SerIleHisMetGlnuLeuLysIleAsnIleSerArgLeuGluIleAspLysGlnIle	2229
Oy	139	-----AAAGAAATTTCTTCTGGCTGCTCAATACTGTGAACCGTGTGT	180
Db	2230	TrpGluSerLysAlaGlnThrGluValGlnIleuGlnGlnLysValCysAspThrLeuGln	2249
Oy	181	GGCTCTTGTGTGAATGTGCTCAGACATGAGAGCACTGCTCTTCCCAAAACCATCTAATGTT	240
Db	2250	Gly-----GluAsnLysGluLeuLeuSerGlnLeuGlnGluIuThrArg	2265
Oy	241	CATATG-----CAGACCATCGAAGAAGCTGTTAAAGAAAGAGATGACTTG	285
Db	2264	HisLeuTyrHisSerSerGlnAsnGluLeuAlaLysLeuGlnuSerGluLeuLysSerLeu	2283
Oy	286	ATGTCGTGCACATGTTCCGTAAAGAGACAGCTTG-----	318
Db	2294	LysAspGlnLeuThrAspLeuSerAsnSerLeuGlnLysCysLysGlnGlnLysGlnLysn	2303
Oy	319	-----GCAGATCCACCAAGAGAGAGACAGTGCCTTAT	351
Db	2304	LeuGlnGlyIleIleArgGlnGlnGlnAlaAspIleGlnAsnSerLysPheSer---Tyr	2322
Oy	352	GACAGAGTGAACACAGATTTCGCAAAATCT-----GAGCA	387
Db	2323	GluIleuLeuGluThrAspLeuGlnAlaSerArgGluLeuThrSerArgLeuHisGlnGlu	2342
Oy	388	GCCAAATTTTGA AAAAACCAAGCGCTTA-----	414
Db	2343	IleAsnMetLysGlnGlnLysIleIleSerLeuLeuSerGlyLysGlnGlnAlaIleGln	2362
Oy	415	ATCCAGTGTGACCAAGTTGAGAGAAAGACTGAGAGAGCGGCGAGCCACTTGAAAAAGAA	474
Db	2363	ValAlaIleAlaGlnLeuArgGlnGlnIlnHisAspLysGluIleLysGlnLeuGlnAsnLeu	2382

C:Date: 19-Oct-1995 #sequence_revision 26-Jan-1996 #text_change 10-Dec-1999
C:Accession: A56539, S37536
R:Seelij, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.
Mol. Cell. Biol. 14, 2564-2576, 1994
A>Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein
A:Reference number: A56539; NUID:94187728; PMID:7511208
A:Accession: A56539
A:Molecule type: mRNA
A:Residues: 1-3259 <S>
A:Cross-references: EMBL:X75304; NID:9405714; PIDN:CAA53052.1; PID:9405715
C:Genetics:
A:Gene: GDB:GOLGB1; GCP: GCP371
A:Cross-references: GDB:454958
A:Map position: 3q13.31-3q13.31
C:Superfamily: giantlin
A:Keywords: coiled coil; Golgi apparatus; transmembrane protein
F:3238-3254/Domain: transmembrane #status predicted <TMN>

Alignment Scores:

Pred. No.	Length:	Matches:
Score:	253.50	119
Percent Similarity:	40.33%	Conservative: 104
Best local Similarity:	21.52%	Mismatches: 179
Query Match:	9.28%	Indels: 151
	Gaps:	24

US-09-502-945-1 (1-1552) x A56539 (1-3259)

```

OY      1 CTTCGTGCATCGACAGCACTGAATAACTTACTTTAT-----GAG   39
Db      2204 VALLLEAPRGUVALVLYSTPTPGULVRGLYSPIHESERRAPLAILEINSErLySGlu 2223
OY      40 GAAGAAGTGTAATTGAGAAATCCCAATTGAAGTTTTTTGGAGAACGACTTAGCTGAATAT  99
Db      2224 GLUGLUILVLRGLYSLVLSPLASPCNSYSSERVALLEULYSAPRGILNLEUARGlImet 2243
OY     100 CAGAGAACTTGTGAAGATCTTAAA-----GAGCACTAAAGCAT-----  138
Db     2244 SErIleHisMetGIUGLULeULysILEnsnIlEsErArGLyEUGLIHisAsPlySgLinIle 2263
OY     139 -----AAGAAATTTCTTGCGTGGCTGTAATCTTGAACCGGTGGT  180
Db     2264 TPrgLuseRLysAlaGlInThhGIVaIGlIneUGlInLysValCYsAsPThLeUGln 2283
OY     181 GGCTTTGTTTGAATGTCGCACAGATGAAGCTGTCTTTGCCAACCCATCTAATGTT  240
Db     2284 GLY-----GLAsnLySglUleUenseRgInLeUGLUthrArg 2297
OY     241 CATATG-----CACACCATCGAAAGACTGGTTAAAGAAAGATGACTTG  285
Db     2298 HisLeuYrHisSerSerGIAnsnglUleUalLysLeUGLuseRgIUleULysSerLeu 2317
OY     286 ATGTCGTGCATAGTTTCCGTAAAGAGCAGCTTG-----  318
Db     2318 LysAsPrGInLeUPThrAsPLeUsErAsnSerIeUDUlYScYsLysGluGlnLySgLIAsn 2337
OY     319 -----GCAGATACGAGCAAAAGACAAGTCTTAT  351
Db     2338 LeUGLUYLleILeARgInGlnGluAlAsPrlEGlInAsenSerLysPheser---Tyr 2356
OY     352 GAACAGGTGAACAAGATTGGCAATATCT-----GAGGAA  387
Db     2357 GlUnleUGlInThrAsPLeUsErGlnAlLseArGLUleUthrSerArGLeUHisSgLUglU 2376
OY     388 GCCAATTTTGAAAAACCAAGCGCTTA-----  414
Db     2377 ILeAsnMeTySgLUgInLysILleISerLeUenseRgLysGluGlnJuaILAIEglIn 2396
OY     415 ATCCAGTGTGCACAGTTGAGAAAGAGCTGGAGAGCAGCGGCGCACTTGAAGAAAGAA  474
Db     2397 VALAlaILEaGlUleUdARgInGlnGlnHisAsPrySglUleYsGluLeUGLAsnLeU 2416
OY     475 CTTCGATCTCAGCAAGAGAAAGGCCATTGAGAAACATGATGAAGAAAGAAATTAACG  534

```


Db 1566 GluGluLeuLeuArgLeuGlnAlaGluGlnAlaGlnArgLeuArgGlnAlaGluValGln 1585
QY 709 ATGGCGTATGAGTGAAT-----AAACCAACATGAGAGATGAGCGCAAGAA 759
Db 1586 ArgAlaArgGlnValGlnValAlaLeuGlnThrAlaGlnArgSerAlaGlnAlaGlnLeu 1605
QY 760 GAGCACAGA-----GAGTTCAGAGCAAAACCTAACAGGATCTTGAATTAAGATCAG 813
Db 1606 GlnSerLysArgAlaSerPheAlaGlnLysThr---AlaGlnLeuGlnArgSerLeuGln 1624
QY 814 GAAATAGAGAAATTGAGATAGAACTGATGAAAGCAACACCTTGAGACGAGACAG 873
Db 1625 GluGlnHisValAlaValAlaGlnLeuArgGlnGlnAlaGlnArgAlaGlnGlnGln 1644
QY 874 CAGAAAGCAGCCCTGCGCAGAGAGAGAGTGCCTGAGCTAACAGAA-----CTGCTG 924
Db 1645 AlaglnAlaGlnArgAlaArgGlnGlnGlnAlaGlnArgGlnLeuGlnGlnGlnLys 1664
QY 925 GGGCAATCTGAGCAGCAACTGACCTCAGCAGATCTGAAATAGCTCAA-----CTC 975
Db 1665 AlaasnGlnAlaLeuArgLeuArgLeuGlnAlaGlnGlnValAlaGlnGlnLysSerLeu 1684
QY 976 AGTCAA---GAAAAAGGTATACATATGATTAATTTGGAAAGTTCACAGAGAAATGAA 1032
Db 1685 AlaglnAlaGlnAlaGlnLysGlnLysGlnLysGlnAlaGlnArgGlnArgGln 1704
QY 1033 GAATTTGAGAGACAGTGTCTCCAGCATGGAGATGACATGACAGATGAAAGAGGCTA 1092
Db 1705 LysAlaGlnGlnGlnAlaValAlaArgGlnArgGlnLeuAlaGlnGlnGlnGlnGln 1724
QY 1093 AGCAGCTGATTAAGACACAGCCAGCCAGCCAGCAG----- 1131
Db 1725 ArgGlnLeu-----AlaGlnGlnLysThrAlaGlnGlnArgLeuAlaAlaGlnGln 1741
QY 1132 CTGTGTCAGCTCCGACAGAGCAAGCAAGCACTCTCTCGAGAGCAGACCTGTGGAA 1191
Db 1742 LeuLeuArgLeuArgAlaGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1761
QY 1192 GAGGTGACCGCTGCGCGAGCCAGTTACCCAGCATGCCAGAA 1233
Db 1762 GluLeuAlaArgLeuGlnArgGlnAlaAlaAlaAlaThrGln 1775
RESULT 12
B43402
myosin heavy chain-B, neuronal - chicken
N:Contains: myosin ATPase (EC 3.6.4.1)
C:Species: Gallus gallus (chicken)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Apr-2002
R:Accession: B43402; A43402
R:Takamashi, M.; Kawamoto, S.; Adelstein, R.S.
J. Biol. Chem. 267, 17864-17871, 1992
A:Title: Evidence for inserted sequences in the head region of nonmuscle myosin specific
yosin.
A:Reference number: A43402; MID:92388144; PMID:1355479
A:Accession: B43402
A:Molecule type: mRNA
A:Residues: 1-2007 <TA>
A:Cross-references: GB:M93676; NID:q212448; PIDN:AAA48988.1; PID:q212449
C:Comment: Alternately spliced segments 1 and 2 are found exclusively in nonmuscle myo
C:Superfamily: myosin heavy chain; myosin motor domain homology
C:Keywords: actin binding; alternative splicing; ATP; coiled coil; hydrolyase; methylated
F:1-2007/Product: myosin heavy chain-B, neuronal #status predicted <MTN>
F:1-211,222-631,653-2007/Product: myosin heavy chain-B, nonmuscle #status predicted <MTN>
F:178-185/Region: nucleotide-binding motif A (P-loop)
F:212-221/Region: alternatively spliced segment 1 #status experimental
F:559-593/Region: actin binding #status predicted

F:632-652/Region: alternatively spliced segment 2 #status experimental
F:692-714/Region: actin binding #status predicted
F:875-2007/Domain: coiled coil #status predicted <COI>
F:875-1315/Region: S2
F:1316-2007/Region: light meromyosin
F:129/Modified site: N6,N6,N6-trimethyllysine (Lys) #status predicted
F:184/Binding site: ATP (Lys) #status predicted
F:732,742/Active site: Cys #status predicted
F:1954/Binding site: phosphate (Thr) (covalent) #status predicted
F:1987/Binding site: phosphate (Ser) (covalent) #status predicted
Alignment Scores:
Pred. No.: 2,52e-08 Length: 2007
Score: 249.50 Matches: 110
Percent Similarity: 46.58% Conservative: 108
Best Local Similarity: 23.50% Mismatches: 161
Query Match: 9.13% Gaps: 89
Db: 1 Gaps: 20
US-09-502-945-1 (1-1552) x B43402 (1-2007)
QY 1 CTCTGATGATGATCCGAGAGCTTAACCTTATGAGCAAAAGTGAATTTGAGGAA 60
Db 1067 TlethrAspleuGlnArgLeuLysLysGlnLysThrArgGlnGlnGlnGln 1086
QY 61 TCCCAATTGAAGTTT-----TTGAGAGACGACTTACCTGATATCAG 102
Db 1087 AlAlsArgLysLeuAspGlnThrThrAspleuLnsGlnLeAlaGlnGln 1106
QY 103 AGAATCTTGACATCTTAAAGACAACTAAAGCATTAACATTTCTTGGCTGCT--- 159
Db 1107 AlaglnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1126
QY 160 -----AATCTTGAACGCTGTGGNGCTT 186
Db 1127 LeuAlaArgLysAspGlnGlnAlaValAlaGlnLysAsnAlaLeuLysVal----- 1143
QY 187 TGTTTGAATGTCTCAGCATGAGCTGTCTTCCCAACCATACTAATGTTCATATG 246
Db 1144 -----lleArgGlnLeuGlnAlaGlnlleAlaGlnLeuGlnGlnAspLeuGlnSer 1160
QY 247 CAGACCATCCAAAG-----CTCGTTAAAGAAAGATACTGATGCTGCATGTT 300
Db 1161 GluLysAlaSerArgAsnLysAlaGlnLysLysArgAspLeuSerGlnGlnGln 1180
QY 301 TCCGTAGAGACAGCTTGCAGATACGCACCAAGAACAGAGTGGTTATGAA---CAG 357
Db 1181 AlaleuLysThrGlnLeuGlnAspThrLeuAspThrThrAlaAlaGlnGlnGlnGln 1200
QY 358 GTGAAACAGTTTGCATATCTGAG-----GAAGCCAAATTTGAAAAACCAAG--- 408
Db 1201 ThrLysArgGlnGlnGlnValAlaGlnLeuLysAlaGlnGlnGlnGlnGlnGln 1220
QY 409 -----GCTTATCCAG-----TGTGACCGATGAGG 435
Db 1221 HisGlnAlaGlnGlnGlnGlnLysGlnArgHisAlaThrAlaLeuGlnGlnSer 1240
QY 436 AAGGAGCTGAGAGAGCGAGCGAGCACTTGAAGAAAGCACTTCATCTCAGCAAGAGAA 495
Db 1241 GluGlnLeuLeu-----GlnAlaLysArgPheLysAlaAsnLeu-----GluLysAsnLys 1257
QY 496 AGGCGCATGAGAAAGACATGATGAAAGAAAGAAATTAACGAAGAAAGGAGTCAATGGCA 555
Db 1258 GlnGlnLysLeuSerAsp-----AsnLysGlnLeuAlaCysGlnValLysValLeuGln 1275
QY 556 TCAGAGTGTGATCTTGTCTCAGATATTTGCCACATGAGGCCAGCGAGGAAAGGTT 615
Db 1276 GlnValLysAlaGlnSerGlnHisLysArgLysLysLysLysLysLysLysLys 1295
QY 616 ACAAGAGAAAGATTTCAGCTATTATCACTAGAGAGAAATTAACAGCCAGCTGCTTTC 675
Db 1296 Thr-----AlaLysValThrGlnGlnGlnArgLeuArgValGlnLeuAlaGln 1311

```

QY 676 CGGGAATGATGATCACAAGGTGTGTGGAGAAATGCCGTATCAGCTGTAATTAACCAAC 735
   ::
Db 1312 Lys-----AlaasnlysluenglnannglnleuLeuaspasnlalser 1324
QY 736 ATGGAGAAAGATGAGCCGAAAAAGGACACAGAGCTTCAGAGCAAAATTAACAGGGAT 795
   ::::::::::::::::::::
Db 1325 SerleuLeuGluGluAlaArgLysSglYleYlserPhe---AlaYlaspAlaIalaser 1343
QY 796 CTGGAATTAAGATCAAGAAATAGGAAATTG-----828
   |||||
Db 1344 LeuGluSerGlnLeuGlnAspThrGlnGlnLeuLeuGlnGluGluThrArgGlnLysLeu 1363
QY 829 -----AGATA---GAACGTGATGAAGCAAAACAACCTTGAACAGAGCAG 873
   |||||
Db 1364 AsnleuSerSerArgLleArgGlnLeuGlnGluGlnLysAsnAsnLeuGlnGlnGln 1383
QY 874 CAGAAAGCAGCCCTGGCCAGA-----GAGGAGTCCCTGAGACTAACAGACTG 921
   ::::::
Db 1384 GluGluGluGluGluAlaArgLysAsnLeuGlnLysGlnMetLeuAlaLeuGlnAlaGln 1403
QY 922 CTGGCGCAATCTGAGACCACTGCACCTCACAGATCTGAATAGTCAACTCAGTCAA 981
   |||
Db 1404 LeuAlaGluAlaLysLysLysValaspaspaspLeuGlyThrIleGluGlyLeuGln 1423
QY 982 GAAAAAAGCTATACATATGATTAATTGGAAAGTTACAGAGAGAAATGAAGATTGGAG 1041
   |||||
Db 1424 AsnLysLysLysLeuLeuLysAspMetGlnSerLeuSerGlnArg-----LeuGln 1440
QY 1042 GAACAGTGTGTCACATGAGAGTACATGAGAGCATGAAGCAAAAGCTTAAGCAGCTG 1101
   |||||
Db 1441 GluLysAlaMetAlaTyrAspLysLeuGlnLysThrLysAsnArgLeuGlnGlnLeu 1460
QY 1102 GAT-----AAGCACGCCAGGCCACAGCCAGCCAGCTGGTGCAGCTC 1143
   |||
Db 1461 AspaspLeuMetValaspLeuAspHisGlnArgGlnIleValSerAsnLeuGlnLys 1480
QY 1144 CTCGACAGACGAAACCACTTCTCTGAGAGAGCAGACCTGTGCTG-----GAA 1191
   |||
Db 1481 GlnLysLysPheAspGlnMetLeuAlaGlnLysAsnIleSerAlaArgTyrAlaGln 1500
QY 1192 GAGGTGACCGCGCTGGGAGCCAG 1215
   |||
Db 1501 GluArgAspArgAlaGluAlaGln 1508

RESULT 13
A39638
Plectin - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A39638; S21876
R:Wiche, G.; Becker, B.; Huber, K.; Welter, G.; Castanon, M.J.; Hauptmann, R.; Stratowa
J. Cell Biol. 114, 83-99, 1991
A:Title: Cloning and sequencing of rat plectin indicates a 466-kD polypeptide chain with
A:Reference number: A39638; MUID:91268156; PMID:2050743
A:Accession: A39638
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-4687 <RIC>
A:Cross-references: EMBL:X59601; NID:91292885; PIDN:CAA42169.1; PID:91561642
C:Superfamily: plectin; alpha-actinin actin-binding domain homology; ribosomal protein S
C:Keywords: cytoskeleton; transmembrane protein
F:6-103/Domain: ribosomal protein S10 homology <RS10>
F:184-399/Domain: alpha-actinin actin-binding domain homology <ACT>

Alignment Scores:
Pred. No.: 2.89e-08 Length: 4687
Score: 248.50 Matches: 113
Percent Similarity: 44.32% Conservative: 82
Best Local Similarity: 25.68% Mismatches: 164
Query Match: 9.10% Indels: 81
DB: 1 Gaps: 19
US-09-502-945-1 (1-1552) x A39638 (1-4687)

```

```

QY 4 CTGATGATCCGAGACGCTAAAACTTACTTATGAGAAAG-----TGCGAA 51
   |||
Db 1675 LeuGlnAlaLeuAspGlnLeuLysLeuGlnAlaGlnGluAlaGluArgTTrpLeuGln 1694
QY 52 ATTGAGAAATCCCAATTGAAGTTTGTGAGGAACGACTTACGTGAATACAGAACTGT 111
   |||
Db 1695 AlaGlnAlaGlnArgAlaArgGlnValGlnValAlaLeuGlnLysThrAlaGlnArgSerAla 1714
QY 112 GAAGATCTTAAGAGCAACTAAGCATTAAGCAATTTCTTGGCTGCTTAATACTTGTAC 171
   |||
Db 1715 Glu---ValGluLeuGlnSerLysArgProSerPhe-----1725
QY 172 CGTGTGTGGTCTTGTGTTGAATGTCTGCACTGAAGCTGTCTTCCCAACCAT 231
   |||
Db 1726 -----AlaGlnLysThrAlaGlnLeuGlnLysValGlyThrLeuGlnGlnLysHis 1740
QY 232 ACTAATGTTCAATAG-----CAGACCATGAAAGCACTGGTTAAGAAAGAGATGACTTG 285
   |||
Db 1741 ValThrValThrGlnLeuArgGlnLysAlaGlnArgAlaGlnGlnGlnAlaGln 1760
QY 286 ATGTCTGCATAGTTTCCGTAAAGAGCAGCTTGGCAGATACGCAAGCAAGAGCAAGT 345
   |||
Db 1761 GluArgAlaArgGlnGluAlaGlnArgGlnLeu-----GluArgTrpGlnLeuLys 1777
QY 346 GCTTATGAAACAGTGAACAAGTTTTCGAAATATCTGAGAAACCAATTGGAANAAC 405
   |||
Db 1778 AlaAsnGlnAlaLeuArgLeuArgLeuGln---AlaGlnGlnValAlaGlnGlnLysSer 1796
QY 406 AAGCTTTAATTCACGTGTGACCACTTGAAGAGAGCTGAGAGCGAGCGAGCACTT 465
   |||
Db 1797 LeuAlaGlnAlaAspAlaGlnLysGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGln 1816
QY 466 GAAAAAGAACTTGCATCTCAGCAGAGAAAGGCCATTGAGAAAGACATGATAAAG 525
   |||
Db 1817 GlyLys-----AlaGlnGlnGlnAlaValaLysGlnArgGlnLeuAlaGln 1832
QY 526 GAAATTAACGAAAGAAAG-----GAGTATACGGATGATCAAGATGTGACTTGTCT 576
   |||
Db 1833 GluLeuGlnLysGlnArgGlnLeuThrGlnGlnGlnGlnGlnGlnGlnGlnGln 1852
QY 577 CAGAAATATGGCCCACTGAGAGCCGAGGTGAA-----609
   |||
Db 1853 GlnGlnLeuLysLeuArgLeuArgAlaGlnThrGlnGlnGlnGlnGlnGlnGln 1872
QY 610 -----AAGTTTCAAGAAAGATTTACGTATTAATCAACTGAGGAA 654
   |||
Db 1873 GluGlnGlnLeuAlaArgLeuGlnHisGlnLysAlaThrAlaAlaThrGlnLysArgGln 1892
QY 655 ATTCAAGCCAGCTGCGCTTCGGGAATGATGTCACAAAGGTTGTGGAATATGCGC 714
   |||
Db 1893 LeuGlnAlaGlnLeuAla-----LysValArgAlaGlnMetGln 1905
QY 715 TATCAGCTGAATTAACCAACATGAGAGAGATGAGCAGAAAGAGCAGACAGAGTTT 774
   |||
Db 1906 ValLeuLeu-----AlaSerLysAlaArgAlaGlnGlnGlnSerArgSerThr 1921
QY 775 AGACCAAAACTAACAGGATCTTGAATTAA-----GATCAG 813
   |||
Db 1922 SerGlnLysSerLysGlnArgLeuGlnLysAlaGlnLysAlaGlnArgPheArgGlnLeuAlaGln 1941
QY 814 GAAATTAAGAAATTAAGAAATGAACTGATGAAAGCAAAAGCACTTGAACAGAGGAG 873
   |||
Db 1942 GluAlaAlaArgLeuArgAlaLeuAlaGlnLysAlaArgArgHisArgGlnLeuAlaGln 1961
QY 874 CAGAAAGCAGCCCTGGCCAGAGAGAGAGTGCCTGAGA---CTAAGCAAGACTCTGGCGGAA 930
   ::
Db 1962 GluAspAlaAlaArgGlnArgAlaGlnLysAlaGlnLysValLeuThrGlnLysLeuAlaAla 1981
QY 931 TCTGAGCAGCACTGCACTG---ACCAAGATCTGAATAGCTCAACTGATCAAGAAAA 987
   |||
Db 1982 IleSerGlnAlaThrArgLeuLysThrGlnAlaGlnIleAla-----LeuLysGlnLys 1999

```



```
OY 1003 A A A T T G G A A G T T A C A G A G A A A T G A G A A T T G G A G G A C A G T G T C C A G C A T G G G 1062
Db 1471 V a l T h r G l n m e t g l n g l n a s n g l n s e r A r g I l e g l u a r g l u l e u ----- 1486
OY 1063 A G A G T A C A T G A G A C G A T G A A G C A A A G G C T A A G C A G C T G G A T A A G C A C A G C C A G C C A C A 1122
Db 1487 ----- G l n g l u l y s l y s l y s g l n h i s l e t h r A r g l l e a s p a s p g l u a r g a s p g l u l e u 1503
OY 1123 G C C C A G C A G C T G G T G C A G C T C C T C A G C A A G C A A C C A G C T T C C T G G A G A G G C A G A G C 1182
Db 1504 L y s l y s g l n l e u t h r g l n l e u g l n g l n h i s g l u c l n s e r S e r t h r g l n l e u l e u ----- 1523
OY 1183 C T G T C G A A G A G G T G A C C G G T G C G A C C A G T T A C C A G A T G C C A C A A T C T G A T T G C 1242
Db 1524 A l a g l n a s n g l u l e u g l u a r g l e u A r g l y s ----- 1533
OY 1243 T G A C C T G A T G A C A G A G T G A A T A A T G A A T T A C A A A G 1284
Db 1534 ----- L y s g l u l e u l y s t y r l y s g l u ----- 1540
```

Search completed: March 21, 2003, 13:04:27
Job time : 70.6459 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:48:47 ; Search time 39.5926 Seconds
(without alignments)
10292.707 Million cell updates/sec

Title: US-09-502-945-1
Perfect score: 2732
Sequence: 1 ctctgcatgcctccgcgaa.....aaatgaacttttaagaaga 1552

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 609376 seqs, 131287123 residues
Total number of hits satisfying chosen parameters: 1218752

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ n2p model -DEV=xld
-Q=/cgn2_1/USPRO/US09502945/runat_14032003-101100_19241/app_query.fasta.1.10979
-DB=ending_Patents_AA_New -QFMT=fastn -SUFFIX=rapn -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945.ecgn.1.1.226 @runat.14032003.101100.19241 -NCPu=6 -ICpu=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WAIT.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7
-YGAPOP=10 -YGAPEXT=0.5 -DELop=6 -DELEXT=7

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCF_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	666	24.4	264	US-60-443-566-4443 Sequence 4443, Ap
2	666	24.4	362	US-60-443-566-4444 Sequence 4444, Ap
3	537	19.7	172	US-10-276-774-1921 Sequence 1821, Ap
4	280	10.2	1294	US-10-282-1224-61292 Sequence 11292, A
5	268.5	9.8	3852	US-09-724-676-68911 Sequence 68911, A
6	268.5	9.8	3852	US-09-724-676-68912 Sequence 68912, A
7	268.5	9.8	3852	US-09-724-676-68913 Sequence 68913, A
8	268.5	9.8	3852	US-09-724-676-68911 Sequence 68911, A
9	268.5	9.8	3852	US-09-724-676-68912 Sequence 68912, A
10	268.5	9.8	3852	US-09-724-676-68913 Sequence 68913, A
11	268.5	9.8	3864	US-09-724-676-68923 Sequence 68923, A

12	268.5	9.8	3864	US-09-724-676-68924	Sequence 68924, A
13	268.5	9.8	3864	US-09-724-676-68925	Sequence 68925, A
14	268.5	9.8	3864	US-09-724-676-68923	Sequence 68923, A
15	268.5	9.8	3864	US-09-724-676-68924	Sequence 68924, A
16	268.5	9.8	3864	US-09-724-676-68925	Sequence 68925, A
17	268.5	9.8	3868	US-09-724-676-68914	Sequence 68914, A
18	268.5	9.8	3868	US-09-724-676-68914	Sequence 68914, A
19	268.5	9.8	3873	US-09-724-676-68926	Sequence 68926, A
20	268.5	9.8	3873	US-09-724-676-68926	Sequence 68926, A
21	268.5	9.8	3898	US-09-724-676-68931	Sequence 68931, A
22	268.5	9.8	3898	US-09-724-676-68931	Sequence 68931, A
23	268.5	9.8	3910	US-09-724-676-68917	Sequence 68917, A
24	268.5	9.8	3910	US-09-724-676-68917	Sequence 68917, A
25	264	9.7	3861	US-09-724-676-68907	Sequence 68907, A
26	264	9.7	3861	US-09-724-676-68908	Sequence 68908, A
27	264	9.7	3861	US-09-724-676-68909	Sequence 68909, A
28	264	9.7	3861	US-09-724-676-68907	Sequence 68907, A
29	264	9.7	3861	US-09-724-676-68908	Sequence 68908, A
30	264	9.7	3861	US-09-724-676-68909	Sequence 68909, A
31	264	9.7	3873	US-09-724-676-68919	Sequence 68919, A
32	264	9.7	3873	US-09-724-676-68920	Sequence 68920, A
33	264	9.7	3873	US-09-724-676-68921	Sequence 68921, A
34	264	9.7	3873	US-09-724-676-68919	Sequence 68919, A
35	264	9.7	3873	US-09-724-676-68920	Sequence 68920, A
36	264	9.7	3873	US-09-724-676-68921	Sequence 68921, A
37	264	9.7	3877	US-09-724-676-68910	Sequence 68910, A
38	264	9.7	3877	US-09-724-676-68910	Sequence 68910, A
39	264	9.7	3889	US-09-724-676-68922	Sequence 68922, A
40	264	9.7	3889	US-09-724-676-68922	Sequence 68922, A
41	264	9.7	3907	US-09-724-676-68928	Sequence 68928, A
42	264	9.7	3907	US-09-724-676-68928	Sequence 68928, A
43	264	9.7	3919	US-09-724-676-68916	Sequence 68916, A
44	264	9.7	3919	US-09-724-676-68916	Sequence 68916, A
45	262	9.6	860	US-10-072-012-838	Sequence 838, App

ALIGNMENTS

```

RESULT 1
US-60-443-566-4443
: Sequence 4443, Application US/60443566
: GENERAL INFORMATION:
: APPLICANT: CARCILL, Michele
: APPLICANT: BEGOVICH, Ann
: TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
: FILE REFERENCE: C1001447
: CURRENT APPLICATION NUMBER: US/60/443, 566
: NUMBER OF SEQ ID NOS: 25102
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 4443
: LENGTH: 264
: TYPE: PRT
: ORGANISM: Homo sapiens
US-60-443-566-4443

Alignment Scores:
Pred. No.: 3.28e-43 Length: 264
Score: 666.00 Matches: 132
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 24.38% Indels: 0
DB: Gaps: 0

US-09-502-945-1 (1-1552) x US-60-443-566-4443 (1-264)
OY 16 GAGAAAGCTAAAGCTTATGAGAAAGCTGAATGGAATCCATGCAAGTT 75
DB 128 GtutysleuylsleuThTyrGluGluIuysGluIuysGluIuysPhe 147
OY 76 TTGAGGAAGCACTAGCATATCAGAGCACTGTGAAGTCTTAAAGCACTAAAG 135
|||||

```



```

Oy 1210 ACCAGTTACCGATGCACAA 1233
Db 165 LysLysLeuGluGlnIleSergIn 172

RESULT 4
US-10-282-122A-61292
; Sequence 61292, Application US/10282122A
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OR INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282.122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61292
; LENGTH: 1294
; TYPE: PRT
; ORGNISM: Legionella pneumophila
US-10-282-122A-61292

Alignment Scores:
Pred. No.: 2,47e-13 Length: 1294
Score: 280.00 Matches: 106
Percent Similarity: 47.82% Conservative: 113
Best Local Similarity: 23.14% Mismatches: 137
Query Match: 10.25% Indels: 102
DB: 6 Gaps: 21

US-09-502-945-1 (1-1552) x US-10-282-122A-61292 (1-1294)
Oy 52 ATGAGGAATCCCAATGAGTTTGGAGGACGCTTAGTGAATATCAGAGAACTTGT 111
Db 686 LeuGluLeuSergInLeuLysLeuGlnLeuGlnGluGlnLysLeuGlnPheGlnLeu 705
Oy 112 GAAGATCTTAAGACCACTAAGCATAAAGAAATTTCTTGGCGCTGAATACTTGTAAC 171
Db 706 LysGluLeuLysGluLysLeu---HisGluGlnThrThrLeuSer-----Lys 720
Oy 172 CGGTGTGGTGGCTTTGTTGAATGNGCTCAGCATGAAGCTGTCTTCCCAACCAT 231

```

```

Db 721 ArgLeuGlu-----GluGluLeuGlnThrGlnLys 731
Oy 232 ACTAATGTTCAATAG--CAGACCATCGAAGACGTGTTAAAGAAAGATGACTTGTATG 288
Db 732 ThrAsnThrHisGlnGluGlnLeuThrThrLeuGlnArgThrLysGluLysSerLeuAlaSer 751
Oy 289 TCTGCATGAGTTTCCGTGAAGGAGCGAGCTGGCAGATAGCAGCAAGAGCAAGCTCT 348
Db 752 SerSerLeuGluSerLeuArgLysGluLeuHisGluLeuAlaLysLysGluArgSerLeu 771
Oy 349 TAT-----GAACAGGTGAACACAGTTTGGCAATATCTGAC----- 384
Db 772 HisLysThrLeuGluGluLysGlnLeuGlnValGlnGlnLeuGlnGlnLeuAlaGlu 791
Oy 385 -----GAAGCCAAATTTGAAAAAAC 405
Db 792 LysGluLysGluAsnLeuAlaLeuLysLysAlaSerLysGlnSergInHisGlnLysSer 811
Oy 406 -----AAGGCTTAATC-----CAGTGTGACAGTTGAGGAGGAGCTG 444
Db 812 LeuAspLysSerAlaIleGlnSerLeuThrSergInLeuAsnGlnLeuLysLeuGlnLeu 831
Oy 445 GAGAGGCGAGCGGAGCGACTTGAA-----AAGAACTTGCATCTCAGCAAGAGAAA 495
Db 832 GlnLysGln---GluThrLeuGlnLeuGlnLeuLysSerLeuArgLysGlnIleGlnGlu 850
Oy 496 AGGCGCATTGAGAAAGACATGATGAAAGAAATACGAAAGAAAGAGGATACATGGGA 555
Db 851 GlnThrLeuValIleGlnLysLeuLysGlnGlnLysGlnLys----- 867
Oy 556 TCAAGATGTTGATCTTCTCAGAAATTTGCCACAGCGGCGCCAGCTGGAAGAAAGTT 615
Db 868 -----SerAsnThrHisGln---GluGluThrIleGluArgIle 879
Oy 616 ACAAGGAAAGATTTACGATTATTAATCACTGGAGGAATTCAAAGCAGACTG----- 669
Db 880 ThrLysGluLysSerLeuAlaSerAlaSerAlaLeuGlnSerLeuAlaGlyGlnIleGly 899
Oy 670 -----GCTTTCGGGAAATGAGATGTCACAAAGGTGTGGAAATCGGC 714
Db 900 LeuThrArgLysAsnGluGluLysGlnLeuLysLeuThrLysGlnValHisSerLeuSer 919
Oy 715 TATCAGCTGATTAACCAACATG-----GAGAGGAGTGAAGCAGAAAG 759
Db 920 GlnGlnLeuGluGlnLysGlnLeuGlnIleArgGluPheGlnLysGlnLeuGlnLys 939
Oy 760 GAGCAGACAG-----GAGTTTCAGACCAAAACTAACAGGAGATCTTGAATTT 804
Db 940 GlnLysArgValGlnGlnSerGlnLysGlnLysAlaSerAlaLysArgThrValAlaSer 959
Oy 805 AAAGATCAGAAATGAGAAATTGGAATAGACTGATGAAAGCAACAAACACTTGGAA 864
Db 960 LeuArgGlnGlnValIleSerAsnLeuLysLeuGlnLeuGlnGlnValIleGln 979
Oy 865 CAGGAGCGAGGAGGCGCCCTGGCCAGAGAG-----GAGTGCCTGAGACTAACA 915
Db 980 GlnLysGlnLysGlnLysSerLeuLysSerLeuGlnGlnSerGlnIleIleAlaLeuGln 999
Oy 916 GAAGTGTGGCGGATGAGCAGCAACCTGCACCTCAGCAAGATGGAATAGCTCAACTG 975
Db 1000 GlnIleIleGlnAspGlnLysArgGlnLeuGlnLysLysLysIleGlnGlnLeu 1019
Oy 976 AGTCAGAGAAAAAGGTATATCATGATTAATTTGGAAAGTTTACAGAGCAATGAAGAA 1035
Db 1020 ValSerAlaAsnGln---GluLeuGlnLys-----GlnAsnGlnSer 1032
Oy 1036 TTGAGAGAACAGCTGTCTCCACATGGGAGAGTACATGAGAGATGAAGAGGCTAAG 1095
Db 1033 LeuSerLysGlnAsnLeuHisAsnLysAsnThrValGlnAspLeuLysLysLysLeuAsn 1052
Oy 1096 CAGCTGTGAT-----AAGCAGACCGCAGGCGCAGACGCTGCTG 1137

```


APPLICANT: CompuGen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 CompuGen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2
SEQ ID NO 68913
LENGTH: 3852
TYPE: PRF
ORGANISM: Homo sapiens
US-09-724-676-68913

Alignment Scores:
Pred. No.: 2,18e-12 Length: 3852
Score: 268.50 Matches: 121
Percent Similarity: 41.23% Conservative: 107
Best Local Similarity: 21.88% Mismatches: 162
Query Match: 9.83% Indels: 163
DB: Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68913 (1-3852)

QY 1 CTCTGATGATCCGAGAGCTTAACTTATGAGAAAGTGTGA---ATTGAG 57
DB 1797 MetLeuAsnIleSerArgLeuGlnAlaValGluLysLeuGlnAlaIleSer 1816
QY 58 GAA-----TCCCAATTGAATTTTGAAGACGAC 87
DB 1817 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGluLeuMetArgGluSer 1836
QY 88 TTAGCTGATATATCAGAGACTTGTGAAGCTTAAAGACACTAAAGCAATTT 147
DB 1837 PheArgGlnGlySerGlnGlnAlaThrGluSerLeuLysGlnGluLeuArgGluArg 1856
QY 148 CTCTGGCTGTATATCTTAAACCGTGTGGTCTGTGTTGAATGTGCTGACAT 207
DB 1857 LeuHisGluLysSerArgAlaArgGlu-----GlnLeuAlaValGluLeuSerLysAla 1874
QY 208 GAAGCTGCTTCTCCCAACCCATCTAATGTCATATGACAGACACTGAAGACTGTT 267
DB 1875 GluGluValIleAspGlyTyrAlaAspGluLysThrLeu-----PheGluArgGlnIle 1892
QY 268 AAGAAAGAGACTGATGTGTGCCTA-----GTTCCGTAAGC 309
DB 1893 GlnGluLysThrAspIleIleAspArgLeuGlnGluLeuLysAlaSerAsnArg 1912
QY 310 AGAGCTGCGAGATACGACGAAGA----- 336
DB 1913 LeuGlnGluLeuGlnAlaGlnGlnGlnGlnIleGlnGluLysArgGluLeuLeuSerArg 1932
QY 337 -----GAGCAAGTCTTATGACAGGTGAAACAGTTTTCAA 375
DB 1933 GlnLysGlnAlaMetLysAlaGlnAlaGlyProValGluGln-----GlnLeuLeuGln 1950
QY 376 ATATCTGAGAGACCAATTTTGAAGAAACCAAGGCTTTAATCCAGTGTGACCAAGTGAAG 435
DB 1951 GluThrGluLysLeuMetLysGluLysLeuGlnValGlnCysGlnAlaGlnLysValArg 1970
QY 436 AAGGACTGAGAGCGAGCGAGCGAGCTTGAAGAAAGCACTTGCACTGCG----- 486
DB 1971 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValIleGlnGlnValSerArg 1990
QY 487 -----CAAGAGAAAGGGCCATTGAGAAACATG----- 516
DB 1991 PheIleGluLeuGlnGlnGlnLysAsnThrGluLeuMetAspLeuArgGlnGlnAsnGln 2010
QY 517 ---ATGAAAAAGGAATTAACGAAAGAGAGAGTACATGGAGTCAAGATGTTG----- 567
DB 2011 AlaLeuGluLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2030
QY 568 -----ATCTTGCTCAGAAATTTGCCCACTGAGGCCCGAGTGGGAAAG 612

DB 2031 GlnHisGluArgAspValPheGlnGlnGlnIleGlnLysLeuGlnGlnLeuLysVal 2050
QY 613 GTTACAAAGGAAAGATTTACGCTATTAATCACTGAGGAAATTCAGACAGCTGCT 672
DB 2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGlu-----GlnLeuAla 2069
QY 673 TCTCGGGAATGATGTACCAAAAGGTGTGAGAAATG-----CCCTATCAGCTG 723
DB 2070 AsnHisLeuLysGlnLysThrAspLysCysSerGluLeuLeuSerLysGlnGlnLeu 2089
QY 724 AATAAAACCAACATGAGAGAGATGAGGAGAAAGACGACAGAGCTTCAGAGCAAAA 783
DB 2090 GlnArgAspIleGlnGlnArgAsnGlu-----GluIleGluLysLeuGlnPheArgValArg 2108
QY 783 ----- 783
DB 2109 GluLeuGlnGlnAlaLeuLeuGlnAspArgLysHisPheGlnAlaValGlnAlaLysPro 2128
QY 783 ----- 783
DB 2129 GluLeuSerLeuGlnValGlnLeuGlnAlaGlnLysArgAspAlaIleAspArgLysGluLys 2148
QY 784 -----ACTAAC-----AGGATCTTGAATTAAGAT 810
DB 2149 GluIleThrAsnLeuGlnGlnGlnLeuGlnPheArgGlnGluLeuGlnAsnLysAsn 2168
QY 811 CAGGAATATAGAAATTTGAGATATAGACTGAT-----GAAACCAAAACACAC 858
DB 2169 GluGlnValGlnGlnLeuHisMetGlnLeuGlnIleGlnLysLysGlnSerThrThrArg 2188
QY 859 TTGACACAGACGACGAGAGAGCCCTGCGCCACAGAGAGAGTGCCTGACACTTAACGAA 918
DB 2189 LeuGlnGlnLeuGlnGlnGlnLysLysLeuPheLysAspAspMetGluLysLeuGluLeu 2208
QY 919 CTGCTGGCGCAATCTGAG-----CACCAACTGCACTGCACACAGATCGAATA 966
DB 2209 AlaIleLysGlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGluLysPhe 2228
QY 967 GCTCACTCAGTCAGCAAAAAAGGTATATCATATGATTAATTGGA-----AAG 1014
DB 2229 AlaGlnIleIleGlnGlnLysGlnValGluIleAspGlnLeuAsnGlnGlnValThrLys 2248
QY 1015 TTACAGAGAGA-----AATGCAATTTGAGCA----- 1044
DB 2249 LeuGlnGlnGlnLysLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2268
QY 1045 -----CAGTGTGTCACAGAT 1059
DB 2269 IleArgAspLeuGlnLysThrGlnIleGlnLysLeuMetSerAspGlnGlnLysValLysArg 2288
QY 1060 GGGAGAGTACATGAGACGATGAAGCAAAAGGCTAAGCGCTGATGAAGCAACGACGCC 1119
DB 2289 AsnArg-----GluGlnGlnIleIleGlnGlnLeuAsnGlnValIleGlnLys 2303
QY 1120 ACAGCCGACGAGCTGTGTCAGCTCTCAGCAGCAGACAGCACTTCTCTGAGAGGCGAG 1179
DB 2304 LeuGlnGlnGlnLeuAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2320
QY 1180 AGCCTGCGGAGAGGATGAGCGCCGCTGCGGACCCAGCTTA 1218
DB 2321 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 8
US-09-724-676A-68911
Sequence 68911, Application US/09724676A
GENERAL INFORMATION:
APPLICANT: CompuGen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 CompuGen
CURRENT APPLICATION NUMBER: US/09/724,676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 68911
; LENGTH: 3852
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-68911

Alignment Scores:
Prod. No.:      2, 18e-12      Length:      3852
Score:          268.50        Matches:      121
Percent Similarity: 41.23%    Conservative: 107
Best Local Similarity: 21.88% Mismatches:     162
Query Match:     9.83%        Indels:       163
DB:              5           Gaps:         22

US-09-502-945-1 (1-1552) x US-09-724-676A-68911 (1-3852)

QY      1  CTCTGGATGCATCCGAGAAGCTAAACTTATGAGAAAGTGTGA--ATTGAG  57
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1797 MetLeuAsnIleSerSerArgLeuGlnAlaValGluLeuGlnAlaIleSer 1816

QY      58  GAA-----TCCCAATTGAAGTTTGGAGAACGAC  87
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1817 GluIthrSerSerGlnLeuGlnHisAlaLysValThrGlnIleuMetArgGluSer 1836

QY      88  TTAGCTGAATATCAGAGAACTTGTGAAGATCTTAAGAGCACTAATAAGAAATTT 147
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1837 PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysGlnGlnGluLeuArgGluArg 1856

QY     148  CTCTGGCTCTAATTACTTGAACCGTGTGGTGTCTTTGTTGAATGCTCAGCAT  207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1857 LeuHisGlnGlnSerArgAlaArgGlu-----GlnLeuAlaValGlnLeuSerLysAla 1874

QY     208  GAAGCTGTCTCTCCCAACCCACTAATGTTCAATGTCAGACACGAAAGACTGGTT  267
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1875 GluGlyAlaIleAspGlyTyrAlaAspLysThrLeu-----PheGlnArgGlnIle 1892

QY     268  AAGAAGAGATGACTGTGATGCTGCACATA-----GTTCCGCTAAG  309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1893 GlnGluLysThrAspIleIleAspArgLeuGlnGlnLeuLeuLysAsnArg 1912

QY     310  AGCAGCTGGCAGTACCGCAGCAAGA-----  336
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1913 LeuGlnGlnLeuGlnAlaGlnGlnGlnGlnIleGlnGlnGluArgGlnLeuLeuSerArg 1932

QY     337  -----GAAGCAAGTCTATGAGAGCGAAGAAAGTTTGC  375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1933 GlnLysGlnAlaMetLysAlaGlnAlaGlyProValGlnGln-----GlnLeuGln 1950

QY     376  ATATCTGAGAGAGCCAAATTTTGAAGAAACCAAGGCTTTAATCCAGTGTGACAGTTGAG  435
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1951 GluThrGlnLysLeuMetLysGlnLysLeuGlnValGlnLysGlnAlaGlnLysValArg 1970

QY     436  AAGAGCTGGAGAGCGCGGAGCCGCTTGAAGAAAGACTTGCATCTCAG-----  486
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1971 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnGlnValSerArg 1990

QY     487  -----CAGAGAAAGAGGCCCATGAGAAAGACAG-----  516
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     1991 PheIleGlnLeuGlnGlnGlnLysAsnThrGlnLeuMetAspLeuArgGlnGlnAsnGln 2010

QY     517  ---ATGAAGAAAGAAATACGAAAGAAAGAGAGTACATGAGATCAAGATGTTG-----  567
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2011 AlaLeuLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2030

QY     568  -----ATCTGTCTCAGAAATATTTGCCCACTGCGAGGCCAGAGTGGAAAG  612
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2031 GlnHisGlnLysArgAspValPheGlnGlnGlnIleGlnLysLeuGlnGlnLysVal 2050

QY     613  GTTACAAAGAAAGATTCAGATTAATCAACTGAGAAATTCAGACAGCTGGCT  672
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGlu-----GlnLeuAla 2069

QY     673  TCTCGGAATGTGATGTACAAAGTGTGTGAGAAATG-----CGCTATCAGCTG  723
```

```
Db     2070 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2089
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY     724  AATTAACCAACATGAGAAAGATGAGCAAGAAAGAGCAGACAGAGTTTCAGAGCAAA  783
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2090 GlnArgAspIleGlnGlnLysArgAsnGln-----GlnIleGlnLysLeuGlnPheArgValArg 2108

QY     783  -----  783
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2109 GluLeuGlnGlnAlaLeuLeuGlnLysAspArgLysHisPheGlnAlaValGlnAlaLysPro 2128

QY     783  -----  783
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2129 GluLeuSerLeuGlnValGlnLeuGlnAlaGlnLysArgAspAlaIleAspArgLysGlnLys 2148

QY     784  -----ACTAAC-----AGGATCTTGAATTAAGAT  810
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2149 GluIleThrAsnLeuGlnGlnGlnGlnLeuGlnGlnPheArgGlnGlnGlnLysAsn 2168

QY     811  CAGGAATATAGAAATATGAGAAATGAGACTGGAT-----GAAGCAAAACACAC  858
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2169 GluGlnValGlnGlnGlnHisMetGlnLeuGlnIleGlnLysLysGlnSerThrThrArg 2188

QY     859  TTGGACAGAGAGCAGCAGAGAGCCCTGCGCCAGAGAGAGTGCCTGAGACTAACAGAA  918
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2189 LeuGlnLysLeuGlnGlnGlnLysAsnLysLeuPheLysAspAspMetGlnLysLeuGlnLys 2208

QY     919  CTGCTGGCGCAATCTGAG-----CACCACTGCACCTCACACAGATCTGAATA  966
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2209 AlaIleLysGlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlyLysPhe 2228

QY     967  GCTCACTCAGTCAAGAAAGATATACATATGATTAATTTGGGA-----AAG  1014
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2229 AlaGlnIleIleGlnGlnLysGlnValGlnIleAspGlnLeuAsnGlnGlnValThrLys 2248

QY     1015  TTACAGAGAGA-----AATGAAGAAATTCAGAGAA-----  1044
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2249 LeuGlnGlnGlnLeuLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2268

QY     1045  -----  1059
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2269 IleArgAspLeuGlnThrGlnIleGlnLysLeuMetSerAspGlnGlnLysValLysArg 2288

QY     1060  GGGAGATACATGAGAGATGAGCAAGCAAGCTAAGCAGCTGATTAAGCAGACAGAGGCC  1119
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2289 AsnArg-----GlnGlnGlnIleGlnGlnLeuAsnGlnValIleGlnLys 2303

QY     1120  ACAAGCCAGAGCTGTGTGACAGCTCTCAGCAAGCAGACAGCTTCTCGAGAGAGCAG  1179
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2304 LeuGlnGlnGlnLeuAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2320

QY     1180  AGCCTGTGGAAGAGTGTGACCGGCTGCGGACCCAGTTA  1218
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db     2321 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 9
US-09-724-676A-68912
; Sequence 68912, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68912
; LENGTH: 3852
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-68912

Alignment Scores:
```

Pred. No.: 2.18e-12 Length: 3852
 Score: 268.50 Matches: 121
 Percent Similarity: 41.23% Conservative: 107
 Best Local Similarity: 21.88% Mismatches: 162
 Query Match: 9.83% Indels: 163
 DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676A-68912 (1-3852)

1 CTTTGATGATCCGAGAGCACTTAATCTTATGAGAAAGTGGA---ATTGAG 57
 1797 MetLeuAsnIleSerSerArgLeuGlnAlaValGluLysLeuGlnAlaIleSer 1816
 58 GAA-----TCCCAATTGAGATTTTGGAGAACGAC 87
 1817 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGluLeuMetArgLys 1836
 88 TTAAGCTGAATTCAGAGAACTTGAGATCTTAAAGACAACTAAAGCATTAAGATTT 147
 1837 PheArgGlnLysGlnLysAlaThrGlnSerLysCysGlnGluLysLeuArgLys 1856
 148 CTTTGCGCTGTAATACTGTGAACGGTGTGGTCTTTGTTGAATGTGCTGACAT 207
 1857 LeuHisGlnLysSerArgAlaArgGlu-----GlnLeuAlaValGluLeuSerLysAla 1874
 208 GAAGCTGTTCTTCCCAAAACCATACTAATGTTCAATGACAGACATGAAGACTGTT 267
 1875 GlnGlyValIleAspGlyLysAlaSerGlnLysThrLeu-----PheGlnArgGlnIle 1892
 268 AAGAAAGAGATGACTGTGATCTGCACATA-----GTTTCCGTAAGC 309
 1893 GlnGlnLysThrAspIleLeuAspArgLeuGlnGlnLeuLysAlaSerAsnArg 1912
 310 ACACGCTTGGCAGATACGACCAAGA----- 336
 1913 LeuGlnGlnLeuGlnAlaGlnIleGlnGlnGlnArgGlnLeuLeuSerArg 1932
 337 -----GAAGCAAGTCTTATGAACAGAGTGAACAACTTTTGCAA 375
 1933 GlnLysGlnAlaMetLysAlaGlnLysAlaGlnProValGlnGln-----GlnLeuLeuGln 1950
 376 ATATCTGAGGAAGCAATTTTGAAGAAACCAAGGCTTTAATCAGTGTGACCAAGTGAAG 435
 1951 GlnThrGlnLysLeuMetLysLysLeuGlnValGlnCysGlnAlaGlnLysValArg 1970
 436 AAGGAGCTGAGAGCAGCGAGCAGCTTGAAGAAAGCACTTGACATCTGAG----- 486
 1971 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnValSerArg 1990
 487 -----CAAGAGAAAGGGCCATTGAGAAAGACATG----- 516
 1991 PheIleGlnLysGlnGlnLysAsnThrGlnLeuMetAspLeuArgGlnLysAsnGln 2010
 517 ---ATGAAAGAAAGAAATAAGAAAGAGGAGTACATGGAATCAAGATGTG----- 567
 2011 AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2030
 568 -----ATCTTGTCTCAGATATTGCCCACTGAGAGCCAGCGTGAAG 612
 2031 GlnHisGlnArgAspValPheGlnGlnGlnIleGlnLysLeuGlnGlnLysVal 2050
 613 GTTACAAAGAAAGATTTACCTATTAATCACTGAGAGAAATTCAAAGCCAGCTGGCT 672
 2051 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGln-----GlnLeuAla 2069
 673 TCTCGGGAATGATGTCACAAAGGTGTGAGAAATG-----CGCATATACGCTG 723
 2070 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2089
 724 AATAAACAACATGAGAGAGCAAGAAAGAGCAGACAGAGTGTGAGACAATA 783
 2090 GlnArgAspIleGlnGlnLysAsnGln-----GlnIleGlnLysLeuGlnPheArgValArg 2108

783 ----- 783
 2109 GluLeuGlnGlnAlaLeuLeuGlnLysAspArgLysHisPheGlnValAlaGlnAlaLysPro 2128
 783 ----- 783
 2129 GluLeuSerLeuGlnValGlnLeuGlnAlaGlnArgAspAlaIleAspArgLysGlnLys 2148
 784 -----ACTAAC-----AGGAGCTTGAAATTAAGAT 810
 2149 GluIleThrAspLeuGlnGlnGlnLeuGlnGlnPheArgGlnGlnLysLeuLysAsn 2168
 811 CAGGAATAGAGAAATTTGGAATAGACTGAT-----GAAACCAACACAC 858
 2169 GlnGlnValGlnGlnLeuHisMetGlnLeuGlnIleGlnLysGlnSerThrThrArg 2188
 859 TTGGAACAGGACACAGAGAGCCCTGGCCACAGAGAGTGCCTGAGCTAACAGAA 918
 2189 LeuGlnGlnLysGlnGlnLysLeuPheLysAspAspMetGlnLysLeuGlnLeu 2208
 919 CTGCTGGCGCAATCTGAG-----CACCAACTGCACCTGACACAGATCTGAATA 966
 2209 AlaIleLysGlnLysSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlnLysPhe 2228
 967 GCTCAACTGATCAGCAAGAAAGGATACATATGATTAATTTGGA-----AAG 1014
 2229 AlaGlnIleIleGlnGlnLysGlnValGlnIleAspGlnLeuAsnGlnValThrLys 2248
 1015 TTACAGACAGAGA-----AATGAAGATTTGAGAA----- 1044
 2249 LeuGlnGlnGlnLysLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2268
 1045 -----CAGTGTGCCAGCAT 1059
 2269 IleArgAspLeuGlnThrGlnIleGlnLysLeuMetSerAspGlnGlnLysValLysArg 2288
 1060 GGGAGATGATCATGAGACGATGAGCAAGAGCTTAAGGACACTGATTAAGACACAGGCC 1119
 2289 AsnArg-----GlnGlnGlnIleGlnGlnLeuAsnGlnValIleGlnLys 2303
 1120 ACAGCCGAGCAGCTGTGCTGCTCCTCAGCAGCAAGAACACACTTCTCTGAGAGGCCAG 1179
 2304 LeuGlnGlnGlnLeuAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2320
 1180 AGCCTGTCGGAAGAGTGGACCGGCTGCGACCCAGTTA 1218
 2321 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 10
 US-09-724-676A-68913 ; Sequence 68913, Application US/09724676A
 ; GENERAL INFORMATION:
 ; APPLICANT: Comugen LTD
 ; TITLE OF INVENTION: Variants of alternative splicing
 ; FILE REFERENCE: 129181.4 Comugen
 ; CURRENT APPLICATION NUMBER: US/09/724.676A
 ; NUMBER OF SEQ ID NOS: 2000-11-28
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO: 68913
 ; LENGTH: 3852
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-724-676A-68913

Alignment Scores:
 Pred. No.: 2.18e-12 Length: 3852
 Score: 268.50 Matches: 121
 Percent Similarity: 41.23% Conservative: 107
 Best Local Similarity: 21.88% Mismatches: 162
 Query Match: 9.83% Indels: 163
 DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676A-68913 (1-3852)

[illegible]

```

QY 784 -----ACTAAC-----AGGAGCTTGAATTTAAAGT 810
Db 2149 GtUlelterAsnLeuGluGluGlnLeuGluGlnPheArgGluGluGlnUlnUlnAsn 2168
QY 811 CAGGAATAATGAGAAATTGAGATATGAGACTGTGAT-----GAAAGCAAAACACAC 858
Db 2169 GtUluValGlnGlnGlnUlnHisMetGlnLeuGluGlnLeuGlnLysLysLysLysLysThrArg 2188
QY 859 TTGGACACAGACGACGACGAAGAGCCCTGGCCAGAGAGAGTGCCTGAGACTAACAGAA 918
Db 2189 LeuGlnGlnLeuGlnGlnGlnUlnAsnLysLeuPheLysAspPheGluLysLeuGlnLeu 2208
QY 919 CTGGGGGGCGAATCTGAG-----CACCACTGCACCTGCACGATCTCAATA 966
Db 2209 AlaLeuLysGlnUlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlyLysPhe 2228
QY 967 GCTCAACTCACTCAAGAAAAAGGTATACATATGATTAATTGGGA-----AAG 1014
Db 2229 AlaGlnLeuLeuGlnGlnUlnLysGlnValGlnUlnLeaSpGlnLeuAsnGlnUlnAlaThrLys 2248
QY 1015 TTACAGACAGA-----AATGAAGATTGGAGAA-----1044
Db 2249 LeuGlnGlnGlnLeuLysLysLeuThrThrAspAsnLysValLeuGluGlnLysAsnGlnLeu 2268
QY 1045 -----CAGTGTCTCAGCAT 1059
Db 2269 IleArgAspLeuGlnUlnThrGlnLeuGlnCysLeuMetSerAspGlnGlnUlnLysValLysArg 2288
QY 1060 GGGAGAGTACATGAGACATGAAAGCAAAAGCTTAAGCAGCTGATTAAGCAGCCAGGCC 1119
Db 2289 AsnArg-----GluGlnGlnUlnLeuGlnUlnAsnGlnUlnValLeuGlnLys 2303
QY 1120 ACAGCCGACGACGCTGGTGACGCTCTCAGAACACAAACAGCTTCTCTGAGAGAGCAG 1179
Db 2304 LeuGlnGlnGlnLeuLeuAlaAsnLysLeuGlnLysThrSer-----MetAsnAlaHis 2320
QY 1180 AGCCTGTGGAAGAGTGAGTGAACCGGCTGCGGAGCCAGCTTA 1218
Db 2321 SerLeuSerGlnGlnUlnAlaAspSerLeuLysHisGlnLeu 2333

RESULT 11
US-09-724-676-68923
; Sequence 68923, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68923
; LENGTH: 3864
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68923

Alignment Scores:
Pred. NO.: 2,18e-12 Length: 3864
Score: 266.50 Matches: 121
Percent Similarity: 41.23% Conservative: 107
Best Local Similarity: 21.88% Mismatches: 162
Query Match: 9.83% Indels: 163
DB: Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68923 (1-3864)
QY 1 CTTCTGGATGATCGGAGAGCTAAACTTACTTATGAGAAAGTGA---ATTGAG 57
Db 1809 MetLeuAsnLysSerSerArgLeuGlnAlaAlaValGlnLysLeuGlnUlnAlaLeuSer 1828
QY 58 GAA-----TCCCAATGAAGCTTTTGAGAACGAC 87

```

```

Db 1829 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetArgLysUser 1848
QY 88 TTAGCTGAATATACAGAACTTGTGAAGATCTTAAGAGCACTAAAGCAATTAAGATTT 147
Db 1849 PheArgGlnLysGlnLysValThrGlnSerLysCysGlnLysGlnLysValArgGlnLys 1868
QY 148 CTTCTGCTGCTATACCTTGTAAACCGTGTGTGCTTTGTTGTAATGTGCTCAAGCAT 207
Db 1869 LeuHisGlnLysSerArgAlaArgGln-----GlnLeuAlaValGlnLeuSerLysAla 1886
QY 208 GAAGCTGTCTTTCCCAACCCACTAATGTTCAATGTCAGACACTCGAAGACTGTGTT 267
Db 1887 GlnGlnValIleAspGlnLysValThrLeu-----PheGlnArgGlnIle 1904
QY 268 AAGAAAGAGATGACTGTGATGCTGCACAT-----GTTTCCGTAAG 309
Db 1905 GlnGlnLysThrAspIleLeuAspArgLeuGlnGlnLeuLeuLysAlaSerAsnArg 1924
QY 310 AGCAGCTTGCGAGATACGCAAGCAAGA----- 336
Db 1925 LeuGlnGlnLysGlnLysValGlnGlnGlnIleGlnGlnLysArgLysLeuLeuSerArg 1944
QY 337 -----GAGCAAGTGCCTTATGACAGGTGAACAAGTTTGCA 375
Db 1945 GlnLysGlnAlaMetLysAlaGlnLysValGlnLysValGln-----GlnLeuLeuGln 1962
QY 376 ATATCTGAGAGAACCAATTTGAAAAAACCAAGCTTAAATCCAGTGCAGCTGAGAG 435
Db 1963 GlnThrGlnLysLeuMetLysGlnLysGlnLysValGlnCysGlnAlaGlnLysValArg 1982
QY 436 AAGGACCTGAGAGGAGCGGAGGAGCACTTGAAGAAAGCACTTCATCTCAG----- 486
Db 1983 AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnLysGlnValSerArg 2002
QY 487 -----CAAGGAAAGAGGCCATGAGAAAGACATG----- 516
Db 2003 PheIleGlnLysGlnLysValAsnThrGlnLeuMetAspLeuArgGlnGlnAsnGln 2022
QY 517 ---ATGAAAGAGAAATTAACGAAAGAGAGAGATACATGAGATCAACAAAGTGTG----- 567
Db 2023 AlaLeuGlnLysGlnLysGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg 2042
QY 568 -----ATCTGTCTCAGAATATGCCCCAATGAGAGGCCAGTGGAAG 612
Db 2043 GlnHisGlnLysArgValPheGlnGlnLysGlnIleGlnLysGlnGlnLysVal 2062
QY 613 GTTACAAAGAAAGATTTCACTATTAATCACTGAGAAATTTCAAGCCAGCTGGCT 672
Db 2063 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGln-----GlnLeuAla 2081
QY 673 TCTCGGGAATGAGATGTCACAAAGTGTGTGAGAAATG-----GCTGTTCAGCTG 723
Db 2082 AsnHisLeuLysGlnLysThrArgLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2101
QY 724 AATAAACCAACATGAGAGATGAGGAGCAAGAAAGAGACACAGATGTTCAAGAGCAAA 783
Db 2102 GlnArgAspIleGlnLysArgGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2120
QY 783 ----- 783
Db 2121 GlnLeuGlnGlnAlaLeuLeuGlnLysAspArgLysHisPheGlnAlaValGlnAlaLysPro 2140
QY 783 ----- 783
Db 2141 GlnLeuSerLeuGlnLysAlaGlnLeuGlnAlaGlnLysArgAspAlaIleAspArgLysGlnLys 2160
QY 784 -----ACTAAC-----AGGATCTTGAATTAAGAT 810
Db 2161 GlnIleThrGlnLysGlnLysGlnLysGlnLysPheArgGlnLysGlnLysLysAsn 2180
QY 811 CAGGAATGAGAAATTTGAGAACTGAGAT-----GAAAGCAAAACAAC 858

```

```

Db 2181 GlnGlnValGlnLysHisMetGlnLeuGlnIleGlnLysGlnLysGlnLysThrArg 2200
QY 859 TTGACAGAGAGCAGCAGAGAGAGCAGCCCTGCGACAGAGAGTGCCTGAGACTAAGCAAA 918
Db 2201 LeuGlnGlnLysGlnLysGlnLysAsnLysLeuPheLysAspMetGlnLysLeuGlnLys 2220
QY 919 CTGCTGGCGGAATCTGAG-----CACCAATGCACTCACCAGATCTGAATA 966
Db 2221 AlaIleLysGlnLysSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlnLysPhe 2240
QY 967 GCTCAACTCAGTCAAGAAAAAGCTATACATGATGAATTTGGA-----AAG 1014
Db 2241 AlaGlnIleIleGlnLysGlnLysValGlnIleAspGlnLeuAsnGlnLysValThrLys 2260
QY 1015 TTACAGAGAA-----AATGAGAAATTTGAGGAA----- 1044
Db 2261 LeuGlnGlnLysLysIleThrThrAspAsnLysValIleGlnLysAsnGlnLeu 2280
QY 1045 -----CAGTGTGTCAGCAT 1059
Db 2281 IleArgAspLeuGlnLysThrGlnIleGlnLysLysMetSerAspGlnLysValLysArg 2300
QY 1060 GGGAGAGTACATACAGACATGAGCAAGCAAGCTAAGCAGCTGATGAACAGCCAGGCC 1119
Db 2301 AsnArg-----GlnGlnGlnIleGlnLysGlnLysGlnLysValIleGlnLys 2315
QY 1120 ACAGCCAGCAGCTGCTGAGCTCTCCAGCAACAGCAACAGCTCTCTGAGAGCGAG 1179
Db 2316 LeuGlnGlnLysLeuAlaAsnIleGlnLysGlnLysThr-----MetAsnAlaHis 2332
QY 1180 AGCTGTGCGAGAGGTGACCGCGCTGCGAGCCACTTA 1218
Db 2333 SerLeuSerGlnLysAlaAspSerLeuLysHisGlnLeu 2345

RESULT 12
US-09-724-676-68924
; Sequence 68924, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68924
; LENGTH: 3864
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-68924

Alignment Scores:
Pred. No.: 2,18e-12 Length: 3864
Score: 268.50 Matches: 121
Percent Similarity: 41.23% Conservative: 107
Best Local Similarity: 21.88% Mismatches: 162
Query Match: 9.83% Indels: 163
DB: Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68924 (1-3864)
QY 1 CTTCTGAGATGATCGAGAGACTTAAGACTTATGAGAGAAAGTGTGAA---ATTGAG 57
Db 1809 MetLeuAsnIleSerSerArgLeuGlnAlaAlaValGlnLysLeuGlnAlaIleSer 1828
QY 58 GAA-----TCCCAATTTGAGTTTGTGAGAGAGCAG 87
Db 1829 GlnThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetArgGlnLys 1848
QY 88 TTAGCTGAATATACAGAACTTGTGAAGATCTTAAGAGCACTAAAGCAATTAAGATTT 147
Db 1849 PheArgGlnLysGlnLysValThrGlnSerLysCysGlnLysGlnLysValArgGlnLys 1868

```


OY	148	CTTCTGGCGTGTAACTACTTGTAAACCGGTGGGCTCTTTGTTGAAATGCTCAGCAT	207
Db	1869	LeuHisgInuSerArgAlaArgGlu-----GlnLeuAlaValGluLeuSerLysAla	1866
OY	208	GAACGTGTCTTCTCCCAAAACCCACTACTAAATGTTTCATATGCACAGCATCGAAAGACTGGATT	267
Db	1887	GluGluValIleAspGlyTyrTrpAlaAspGluLysThrLeu-----PheGluArgGlnIle	1904
OY	268	AAAGAAAGACATGACTTGATGTCTGCAGTA-----GTTTCCGTAAAG	309
Db	1905	GlnGlnLysThrAspIleIleAspArgLeuGlnGlnIleuLeuLysAlaSerAsnArg	1924
OY	310	AGCAGCTGGCAGATACGACGAAAGA-----	336
Db	1925	LeuGlnLeuLeuGlnAlaGlnGlnGlnGlnIleGlnGlnLysArgGluLeuLeuSerArg	1944
OY	337	-----GAAGCAAGTGGTTATGACACAGGTGAACCAAGTTTGGCA	375
Db	1945	GlnLysGlnAlaMetLysAlaGlnAlaGlyProValGlnGln-----GlnLeuLeuGln	1963
OY	376	ATATCTGAGGAAGCCAAATTTTGAAAAAACCAAGGCTTTAATCCAGTGTGACCAAGTTTGAG	435
Db	1963	GluThrGlnLysLeuMetLysGlnLysLeuGlnValGlnCysGlnAlaGluLysValArg	1982
OY	436	AAGAGCTGGAGAGGCGAGCGGCGAGCCACTTGGAAAAAGAACTTGCACTTCAG-----	486
Db	1983	AspAspLeuGlnLysGlnValLysAlaLeuGlnIleAspValGlnGlnValLysSerArg	2002
OY	487	-----CAAGGAAAGAGGGCCATGTGAAGAAACATG-----	516
Db	2003	PheIleGluLeuGlnGlnGlnLysAsnThrGlnLeuMetAspLeuArgGlnIleAsnGln	2022
OY	517	---ATGAAAAAGAAATTAACGAAAGAAAGGAGATACATGGATCAAAAGATGTTG-----	567
Db	2023	AlaLeuGlnLysGlnLeuGlnLysMetArgLysPheLeuAspGlnGlnAlaIleAspArg	2042
OY	568	-----ATCTTGCTCAGAAATATTGTCCCAACTGGAGCCAGCTGGAAAG	612
Db	2043	GlnHisGlnArgAspValPheGlnGlnIleGlnLysLeuGlnGlnIleuLysVal	2062
OY	613	GTTACAAAGGAAAGATTTCAGCTATTATTAATCACTGAGCAATTCAAAGCCAGCTGGCT	672
Db	2063	ValProArgPheGlnProIleSerGlnHisGlnThrArgGluValGln---GlnLeuAla	2081
OY	673	TCTCGGAAATGATGTCAACAAGGCTGTGTGGAAATG-----CGTATCAGCTG	723
Db	2082	AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuLeuSerLysGlnIleu	2101
OY	724	AATTAACCAACATGAGAGAGATGAGGCCAGAAAGAGCAGACAGAGTTTCAGACCAAA	783
Db	2102	GlnArgAspIleGlnGlnArgAsnGln---GluIleGlnLysLeuGlnPheArgValArg	2120
OY	783	-----	783
Db	2121	GlnLeuGlnAlaLeuLeuGlnLysAspArgLysHisPheGlnAlaValGlnAlaLysPro	2140
OY	783	-----	783
Db	2141	GluLeuSerLeuGlnValGlnLeuGlnAlaGlnArgAspAlaIleAspArgLysGlnLys	2166
OY	784	-----ACTAAC-----AGGAGCTTGAATTAAGAT	810
Db	2161	GluIleThrAsnLeuGlnGlnGlnIleuGlnIleuArgGlnGlnIleuLysAsn	2180
OY	811	CAGGAATTAAGAAATTGAGAAATGAAGCTGGAT-----GAAAGCAACAAACAC	858
Db	2181	GlnGlnValGlnGlnLeuHisMetGlnLeuGlnIleGlnLysLysGlnSerThrThrArg	2200
OY	859	TTGGAACAGAGCAGCAAGAGCCAGCCCTGGCCAGACAGAGATGGCTGAGATCAACAA	918
Db	2201	LeuGlnGlnLeuGlnGlnGlnLysAsnLysLeuPheLysAspAspMetGlnLysLeuGlnLys	2220
OY	919	CTGTGGGCGAATCTGAG-----CACCAATCAGCTCACCAGATCTGAATA	966

```

Db 2221 AAlaIlysgUsErAspAlaMeTserThrGlnAspGlnHisValLeuPheGlyLysPhe 2240
      ::::: |||||::: ::::: :::::
Qy 967 GCTCAACCACTGATCAGAAAAAGGTATACATATGATAAATTGGGA-----AAG 1014
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 2241 ALaGlnIleIleGlnGlnLysGlnValGlnIleAspGlnLeuAsnGlnValThrLys 2260
      |||||::: |||||::: |||||::: |||||::: |||||:::
Qy 1015 TTACAGACAGACA-----ATGAAGAAATTGGAGGA----- 1044
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 2261 LeuGlnGlnGlnLeuLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2280
      ::::: ::::: ::::: ::::: :::::
Qy 1045 -----CAGGTGTCCAGCAT 1059
      ::::: ::::: ::::: ::::: :::::
Db 2281 IleArgAspLeuGlnThrGlnIleGlnCysLeuMetSerAspGlnIleCysValLysArg 2300
      ::::: ::::: ::::: ::::: :::::
Qy 1060 GGGAGAGTACATGAGACGATGAAGCAAAAGGCTAAGCAGCTGATTAAGCAGCAGCC 1119
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 2301 AsnArg-----GlnGlnGlnIleGlnGlnLeuAsnGlnValIleGlnLys 2315
      ::::: ::::: ::::: ::::: :::::
Qy 1120 ACAGCCAGCAGCTGTGGACGCTCTCCAGACAGCAGACACAGCTTCTCTGAGAGCGAG 1179
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 2316 LeuGlnGlnGlnLeuAlaAsnIleGlnGlnLysThrSer-----MetAsnAlaHis 2332
      ::::: ::::: ::::: ::::: :::::
Qy 1180 AGCGTGTGGAGAGGTGGACCGCGGTGGCGAGCCAGCCACTTA 1218
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 2333 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2345

RESULT 13
US-09-724-676-68925
; Sequence 68925, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 68925
; LENGTH: 3864
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-68925

Alignment Scores:
Pred. No.: 2,18e-12 Length: 3864
Score: 268.50 Matches: 121
Percent Similarity: 41.23% Conservative: 107
Best Local Similarity: 21.88% Mismatches: 162
Query Match: 9.83% Indels: 163
DB: 5 Gaps: 22

US-09-502-945-1 (1-1552) x US-09-724-676-68925 (1-3864)
Qy 1 CTTGTGGATGCATCCGAGAACTTAACCTTACTTATGAGAAAGTGGAA---ATTGAG 57
      ::|||::: ||| ::|||::: |||::: ||| |||
Db 1809 MetLeuAsnIleSerSerArgLeuGlnAlaIleValGlnLysLeuGlnAlaIleSer 1828
      ::|||::: ||| ::|||::: |||::: ||| |||
Qy 58 GAA-----TCCCATTTAAAGTTTTGAGGAGACAC 87
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1829 GluThrSerSerGlnLeuGlnHisAlaLysValThrGlnThrGlnLeuMetArgLys 1848
      ::|||::: ||| ::|||::: |||::: ||| |||
Qy 88 TTACTGCAATTCAGAGAACTGTGAAGATCTTAAAGACACACTAAACATTAAGAAATTT 147
      ::|||::: ||| |||||::: ||| ::|||:::
Db 1849 PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysCysGlnGlnLysLeuArgGlnArg 1868
      ::|||::: ||| |||||::: ||| ::|||:::
Qy 148 CTTGTGCGCTGAATTACTTTAAACCGTGTGGTGCTTTTGTGTAAGATGTGCTCAGCAT 207
      |||||::: ||| |||||::: ||| ::|||:::
Db 1869 LeuHisGlnLysSerArgAlaArgGln-----GlnLeuAlaValGlnLeuSerLysAla 1886
      ::|||::: ||| |||||::: ||| ::|||:::
Qy 208 GAACGCTGTCTTCCCAAAACCCACTACTAATGTTCAATGCAACACATGGAAGACTGGTT 267
      |||||::: |||||::: |||||::: |||||::: |||||:::
Db 1887 GlnGlnValIleAspGlnLysArgLysValArgLysThrLeu-----PheGlnArgGlnLe 1904

```

[illegible]

QY	1015	TTACAGAGGAAA-----	AAAGGAAGATTGGAGAA-----	1044
		:	:	
Db	2261	LeuGInGInGInLeuLysIleThrThrAspAsnLysValIleGluLysAsnGluLeu		2286
QY	1045	-----	CAAGTGTCCAGCAT	1059
Db	2281	IleArgAspLeuGluThrGlnIleGluLysLeuMetSerAspGlnGluLysValLysArg		2300
QY	1060	GGGAGAGACTGATGAGACGATGAAGCAAGGCTTAAGCAGCTGGATGATGACACAGCCAGCC		1119
Db	2301	AsnArg-----	GInGluGluIleGluGluLeuAsnGluValIleGluLys	2315
QY	1120	ACAGCCAGCAGCTGTGTGAGCTCTCCAGCAACAACACAGCTTCTCCGTGAGAGAGAG		1179
Db	2316	LeuGInGInGluLeuAlaAsnIleGluGlnLysThrSer-----	MetAsnLHis	2333
QY	1180	AGCCTGTGGAGAGGTGACCCGGCTCCGGACCCAGTTA		1218
Db	2333	SerLeuSerGlnGluAlaAspSerLeuLysIleGlnLeu		2345
RESULT 14				
US-09-724-676A-68923				
Sequence 68923, Application US/09724676A				
GENERAL INFORMATION:				
APPLICANT: Compugen LTD				
TITLE OF INVENTION: Variants of alternative splicing				
FILE REFERENCE: 129181.4 Compugen				
CURRENT APPLICATION NUMBER: US/09/724,676A				
CURRENT FILING DATE: 2000-11-28				
NUMBER OF SEQ ID NOS: 97222				
SOFTWARE: PatentIn version 3.2				
SEQ ID NO 68923				
LENGTH: 3864				
TYPE: PRF				
ORGANISM: Homo sapiens				
US-09-724-676A-68923				
Alignment Scores:				
Pred. No.:	2,18e-12	Length:	3864	
Score:	268.50	Matches:	121	
Percent Similarity:	41.23%	Conservative:	107	
Best Local Similarity:	21.88%	Mismatches:	162	
Query Match:	9.83%	Indels:	163	
DB:	5	Gaps:	22	
US-09-502-945-1 (1-1552) x US-09-724-676A-68923 (1-3864)				
QY	1	CTTGTGATGCATCCGAGAGCTAAACCTACTTATGAGGAAAGTGTGA--ATGAG		57
		:	:	
Db	1809	MetLeuAsnIleSerSerArgLeuGlnAlaAlaValGlnLysLeuGlnAlaIleSer		1828
QY	58	GAA-----	TCCCAATTGAAGTTTTGGAGAACAC	87
Db	1829	GluThrSerSerGlnLeuGlnIleAlaLysValThrGlnThrGluLeuMetArgLysSer		1848
QY	88	TTAGCTGAATATCAGAGACTGTGAGAGCTTTAAAGACCAACTAAAGCATTAAGACTTT		147
		:::	:::	
Db	1849	PheArgGlnLysGlnGlnAlaThrGlnSerLeuLysCysGlnGlnGluLeuArgLysArg		1868
QY	148	CTTGTGCTGTAACTACTTTAAACCGTGTGGTGGCTCTTTGTAATGTGCTACGAT		207
Db	1869	LeuHisGlnLysSerArgAlaArgGlu-----	GlnLeuAlaValGlnLeuSerLysAla	1886
QY	208	GAACCTGTCTTTCCCAACACCTACTAATGTTCATATGACACACATGGAAGACTGTT		267
Db	1887	GlnIleValIleAspGlyTyrAlaAspGlnLysThrLeu-----	PheGlnArgGlnIle	1904
QY	268	AAAGAAAGAGATGACTGATGTCTGCACATA-----	GTTCCGCTAAG	309
		:::	:::	
Db	1905	GlnLysLysThrAspIleIleAspArgLeuGlnGlnLeuLeuLysAlaSerAsnArg		1924
QY	310	AGCAGCTTGGCAGATGACGACAAAGA-----		336

Db 1963 gUthGluLysLeuMetLysGluLysLeuGluValGlnCysGlnAlaGluLysValArg 1982
QY 436 AAGGAGCTGAGAGCGGAGCGAGCTTGAAAGAACTTGACATCTGAG----- 486
Db 1983 ASPASPLEuGlnLysGlnValLysAlaLeuGlnLeuAspValGlnValSerArg 2002
QY 487 -----CAAGAGAAAGGGCCATTGAGAAAGACATG----- 516
Db 2003 PheLeuGlnLeuGlnGlnGlnLysAsnThrGlnLeuMetAspLeuArgGlnGlnAsnGln 2022
QY 517 ----ATGAAAAGGAATTAAGAAAGAGGAGTACATGGATCAAGATGTTG----- 567
Db 2023 AlaLeuGlnLysGlnLeuLysMetArgLysPheLeuAspGlnGlnAlaLeuAspArg 2042
QY 568 -----ATCTGTCTCAGAAATATGCCCCAACGTGAGGCCAGCTGGAAG 612
Db 2043 GlnHisGlnArgAspValPheGlnGlnGlnGlnLeuGlnLysLeuGlnGlnLeuVal 2062
QY 613 GTTACAAGAAAGATTTACCTATTAATCAATGAGAAATTCAAAGCCAGTGGCT 672
Db 2063 ValProArgPheGlnProIleSerGlnHisGlnThrArgGlnValGlu--GlnLeuAla 2081
QY 673 TCTCGGGAAATGGATGTCACAAAGTGTGGAGAAATG-----CGCTATCAGCTG 723
Db 2082 AsnHisLeuLysGlnLysThrAspLysCysSerGlnLeuLeuSerLysGlnGlnLeu 2101
QY 724 AATTAACCAACATGAGAGAGATGAGCAGAAAGCACAGAGCTTCAGAGCAAAA 783
Db 2102 GlnArgAspIleGlnGlnArgAsnGlu--GlnIleGlnLysLeuGlnPheArgValArg 2120
QY 783 ----- 783
Db 2121 GlnLeuGlnAlaLeuLeuGlnAspArgLysHisPheGlnValAlaGlnAlaLysPro 2140
QY 783 ----- 783
Db 2141 GlnLeuSerLeuGlnValGlnLeuGlnAlaGlnArgAspAlaIleAspArgLysGlnLys 2160
QY 784 -----ACTAAC-----AGGATCTGAAATTAAGAT 810
Db 2161 GlnIleThrAsnLeuGlnGlnGlnLeuGlnGlnPheArgGlnGlnLeuGlnLysAsn 2180
QY 811 CAGGAAATAGAGAAATTGAAATAGAACTGAT-----GAAAGCAAAACAAAC 858
Db 2181 GlnGlnValGlnGlnLeuHisMetGlnLeuGlnIleGlnLysLysLeuSerThrThrArg 2200
QY 859 TTGGAAACAGAGCAGAGAGAGCCCTGGCCAGAGAGAGTCCCTGAGACTAACAGAA 918
Db 2201 LeuGlnGlnLeuGlnGlnLysAsnLysLeuPheLysAspAspMetGlnLysLeuGlnLeu 2220
QY 919 CTGCTGGCGCAATCTGAG-----CACCAACTGCACCTCACAGATCTGAATA 966
Db 2221 AlaIleLysGlnSerAspAlaMetSerThrGlnAspGlnHisValLeuPheGlnLysPhe 2240
QY 967 GCTCAACTCAGTCAGAGAAAGGTAATGATTAATGAGGGA-----AAG 1014
Db 2241 AlaGlnIleIleGlnGlnLysGlnValGlnIleAspGlnLeuAsnGlnValThrLys 2260
QY 1015 TTACGAGAGAGA-----AATGAAGAATTGAGAGAA----- 1044
Db 2261 LeuGlnGlnGlnLeuLysIleThrThrAspAsnLysValIleGlnGlnLysAsnGlnLeu 2280
QY 1045 -----CAGTGTCTCCAGCAT 1059
Db 2281 IleArgAspLeuGlnThrGlnIleGlnCysLeuMetSerAspGlnGlnCysValLysArg 2300
QY 1060 GGGAGAGTACATAGACAGATGAGCAAAAGCTTAAGGCACTGATTAAGCACAGCCAGGCC 1119
Db 2301 AsnArg-----GlnGlnGlnIleGlnGlnLeuAsnGlnValIleGlnLys 2315
QY 1120 ACAGCCCAAGACGTGTCAGCTCCTCAGACAGAACACACAGCTTCTCTGAGAGAGGAG 1179

Db 2316 LeuGlnGlnGlnLeuAlaAsnIleGlnLysThrSer-----MetAsnAlaHis 2332
QY 1180 AGCCTGTCGAAAGAGGTGGAGCCGGCTGCGGACCCAGTTA 1218
Db 2333 SerLeuSerGlnGlnAlaAspSerLeuLysHisGlnLeu 2345

Search completed: March 21, 2003, 13:52:22
Job time : 90.5926 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:44:22 ; Search time 164.803 Seconds

(Without alignments)

12143.272 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732

Sequence: 1 cttctgtagcgcgcgagaa.....aaatgaacttttaagaaga 1552

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4569144 segs, 644733110 residues

Total number of hits satisfying chosen parameters: 9138288

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL-frame+ n2p model -DEV-xyzp
-Q/cgn2_1/USPRO.spool/US09502945/runatc_14032003_101100_19180/app-query.fasta.1.10979
-DB-pending_Patents_AA.Main -QMT-fastan -SUFFIX-gram -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosun62 -TRANS-human40.cdi
-LIST=45 -DOCALLIG=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE-LOCAL -OUTEXT=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR-US09502945 @cgn.1.1.1769 @runatc_14032003_101100_19180 -NCPu=3
-NO_XLPRY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WAIN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2_6/ptodata/1/paa/PCTUS.COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US06.COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US07.COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US08.COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US081.COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US082.COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US083.COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084.COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085.COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086.COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087.COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088.COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089.COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090.COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091.COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092.COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093.COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094.COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095.COMB.pep.*
20: /cgn2_6/ptodata/1/paa/US096.COMB.pep.*
21: /cgn2_6/ptodata/1/paa/US097.COMB.pep.*
22: /cgn2_6/ptodata/1/paa/US098.COMB.pep.*
23: /cgn2_6/ptodata/1/paa/US099.COMB.pep.*
24: /cgn2_6/ptodata/1/paa/US100.COMB.pep.*
25: /cgn2_6/ptodata/1/paa/US101.COMB.pep.*
26: /cgn2_6/ptodata/1/paa/US102.COMB.pep.*
27: /cgn2_6/ptodata/1/paa/US506.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1990.5	72.9	615	14	US-09-012-141-8
2	1990.5	72.9	615	19	US-09-502-967-70
3	1987.5	72.7	713	24	US-10-031-915-52
4	1296.5	47.5	355	1	PCT-US00-35017A-801
5	971	35.5	203	11	US-08-788-789-12
6	841	30.8	393	27	US-60-389-987-766
7	841	30.8	393	27	US-60-412-418-766
8	537	19.7	172	1	PCT-US01-03800A-1921
9	452	16.5	138	22	US-09-834-366-14401
10	452	16.5	138	27	US-60-197-873-14401
11	449	16.3	94	1	PCT-US02-09921-968
12	310	11.3	61	22	US-09-834-366-21271
13	310	11.3	61	27	US-60-197-873-21271
14	300	11.0	62	20	US-09-621-976-5141
15	300	11.0	62	27	US-60-147-499-5141
16	268	9.8	3899	25	US-10-171-311-4
17	268	9.8	3917	25	US-10-171-311-8
18	264	9.7	3878	25	US-09-614-259-11
19	264	9.7	3907	25	US-10-171-311-2
20	264	9.7	3911	27	US-60-389-987-1839
21	264	9.7	3911	27	US-60-412-418-1839
22	264	9.7	3925	25	US-10-171-311-6
23	262	9.6	860	24	US-10-037-417-59
24	262	9.6	860	24	US-10-080-334-166
25	256.5	9.4	1790	27	US-60-360-039-1586
26	255.5	9.4	576775	12	US-08-895-611-2
27	255.5	9.4	576775	18	US-08-895-611D-2
28	255.5	9.4	576775	12	US-09-458-180-2
29	255.5	9.4	576775	22	US-09-895-611D-2
30	254	9.3	2056	20	US-09-614-150-4824
31	254	9.3	2056	27	US-60-191-637-4824
32	254	9.3	2056	27	US-60-191-637-4824
33	253.5	9.3	3225	27	US-60-389-987-254
34	253.5	9.3	3225	27	US-60-412-418-254
35	253.5	9.3	3244	1	PCT-US01-14827-10091
36	251	9.2	1020	25	US-10-179-131-5887
37	251	9.2	1881	1	PCT-US02-03987-15590
38	251	9.2	1881	24	US-10-032-585-7646
39	251	9.2	1881	24	US-10-072-851-15590
40	251	9.2	1881	27	US-60-314-050-7646
41	249	9.1	1161	27	US-60-173-464-14216
42	249	9.1	1294	20	US-09-614-150-17298
43	249	9.1	1294	27	US-60-191-637-17298
44	249	9.1	1294	27	US-60-191-681-13727
45	246.5	9.0	1355	24	US-10-080-334-161

ALIGNMENTS

RESULT 1
US-09-012-141-8
Sequence 8, Application US/09012141
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Racine, Lisa A.
APPLICANT: Metberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 33

```

CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,141
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ. ID NO.: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 615 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-012-141-8

Alignment Scores:
Pred. No.: 1.79e-137 Length: 615
Score: 1990.50 Matches: 409
Percent Similarity: 83.81% Conservative: 0
Best Local Similarity: 83.81% Mismatches: 0
Query Match: 72.86% Indels: 79
DB: 14 Gaps: 1

US-09-502-945-1 (1-1552) x US-09-012-141-8 (1-615)
QY 16 GGAACCTAAACTTACTATAGGAAAGTGAATGAGAAATCCCAATTGAAGTTT 75
DB 128 GLLVLSLEULSLEUTHRYGLUGLULYSCYSGLUILEGLINUSERGINLEULYSPE 147
QY 76 TTGAGAACGACTTACCTAATATCAGAGAACTTGTGAAGATCTTAAGAGCACTAAAG 135
DB 148 LEUARGASINASPLEUNLEGLIUTRYGINARGTHTHCYSLUASPLEUNLSGLINLEULYS 167
QY 136 CATAAAGAAATTTCTTGGCTGCTAATTAAGTGAACCGTGTGGTCTTTGTTGAAA 195
DB 168 HILSYGLUPHEULEUNLEUAAIASNTHRCYSASNARGVALGLYGLYLEUCYSLEULYS 187
QY 196 TTGGCTCACATGAGACGCTGCTTCCCAACCCATACATTAATCTTATATGAGACATC 255
DB 188 CYSALAGLNIHISGLUAIIVALLLEUSERGLINTHTHISHTASNVAHISMETGLINTHTLE 207
QY 256 GAAAGACTGTTAAAGAAAGAGATGACTGATGCTGCATAGTTCCGTAAGAGACAC 315
DB 208 GUARGLEUVALYLSGLIURGASPRASPLEUMESERIALALEUVALISERIALARGSER 227
QY 316 TTGGCAGATACGACGAAAGAGAACGACGCTTATGAACGAGTGAACAGTTTTCAA 375
DB 228 LEUAIASPTHTGGLNGLNARGIUALASERIALATYRGLUGLIVALLYSGLINVALEUGLN 247
QY 376 AATATCGAGAGACCAATTTTGAAGAAACCAAGGCTTTATATCCAGTGTACCAAGTTGAG 435
DB 248 ILESERGLUGLUALASINPHEGLIULYSTRHLYSALALEULLEGLNYSASPGINLEUADY 267
QY 436 AAGAGACTGAGAGGACGCGAGCAGCTTGAAGAAAGAACTTGATCATCAGCAAGAGAAA 495
DB 268 LYSGLULEUGLIURGLINLAGIUAARGLEUGLULYSGIULEUAIASERGLNGLNGLULYS 287

```

```

QY 496 AGGCCATTGAGAAAGACATGATGAAAAAGAAATAAGAAAGCAAGTACATGGGA 555
DB 288 ARGAIILEGLIULYSASPMETMETLYSGUILETHRYSGLUARGIUTYMETGLY 307
QY 556 TCAAGATGTTGATCTTCTGTCAGAAATATGTCCTGAGGCGCCAGGTGGAAGGTT 615
DB 308 SERLYSMETLEULEUSERGLINASNILLEAGLINLEUGIUALAGIIVALLYSVAL 327
QY 616 ACAAGGAAAAAGATTTACGATTAAATCAACTGAGAGAAATTAAGCCAGCTGGCTTC 675
DB 328 THRYSGLIULYSILLESERIALIIEASNGLINLEUGLULIIEGLINSERGINLEUALASER 347
QY 676 CGGGAATGATGTCACAAAGTGTGTGGAAGAAATGCGCTATGACTGTAATAACCAAC 735
DB 348 ARGILMETASPYALTHTRYVALCYSGIYGLIMETARGTYRGLINLEUNLSYTHRASN 367
QY 736 ATGAGACAGATGACGACAGAAAGAGCCACGAGCTGCAGCAAAACTACAGGAT 795
DB 368 METGLULYSASPIUALAGIULYSGLUHNLSARGGLUPHEARGALALYSTHRSNATGASP 387
QY 796 CTGGAATTTAAGATCAGCAAAATAGAGAAATTAAGAACTGAGTGAAGCAAAACA 855
DB 388 LEUGLUILEYSASPIINGLUILEGLIULYSLEUARGILEGLIUEUNASPIUSERTYSGIN 407
QY 856 CACTTGAACAGAGCAGCAGAGAGCCCTGGCCAGAGAGAGTGTGAGACTTACA 915
DB 408 HISLEUGLUGLINGLINGLINSALIALALEUALARGGLIULYSLEUARGLEUTHR 427
QY 916 GAACCTGCTGGCGCAATCTAGCCCAACTGCACCTC----- 951
DB 428 GLULEUENGLYGLUSERGLIHNISGLINLEUHNISLEUTHRARGINGLULYASPSERILE 447
QY 951 ----- 951
DB 448 GLINGLINSERPHESERTYSGIUALALYSALAGLIALALEUCLINLAGINGLINARGGLUGLN 467
QY 951 ----- 951
DB 468 GLULEUTHRLINLSILEGLINGLIMETGLUALAGLINHISASPLYSTRHGLUANSGLINGLN 487
QY 951 ----- 951
DB 488 TYRLEUENLEUTHSERGLINASNTHRPHLEUETHRLYSLEULYSGLULYSCYSTHRC 507
QY 952 ----- ACCAGATCTGAAATAGCTCAACTCAGT 978
DB 508 LEUAIALYSLEUGLUGLINIIESERGLINLYSTHTRARGSERGLULEALAGINLEUSER 527
QY 979 CAAGAAAAAGATATCATATGATAAATTTGGCAAAAGTTACAGAGAAAGAAATGAAGATTG 1038
DB 528 GINGLULYSARGTYRTHRYTRASPLYLSLEUGLYLYSLEUGLINARGARGASNGIUGLITL 547
QY 1039 GAGGACAGTGTGTCCAGCATGAGAGTACATGACATGACATGAGCAAGCAAGGCTTAAGCCAG 1098
DB 548 GLIUGLINCYSVALINISGLIARGVALHISGLITHTMETLYSLINARGLEUADYGGIN 567
QY 1099 CTGGATTAACACACAGCCAGGCCACAGCCACAGCTGTGTCAGCTCTCTCAGCAAGCAAGAC 1158
DB 568 LEUASPLYHNISERGLINLATHRLAAGLINGLINLEUVALGINLEUENLEUSERLYSGINASN 587
QY 1159 CAGCTTCTCTGAGAGGACGACAGACCTGTGGAAGAGGTGACCGGCTGCGAGCCAGTTA 1218
DB 588 GINLEULENGLIULARGLINSERLEUSERGLIULVALASPRYGLUARGTYRTHRGINLEU 607
QY 1219 CCCAGATGCCCAATCTGATTGC 1242
DB 608 PROSERMETPROGLINSERASPCYS 615

RESULT 2
US-09-502-967-70
; Sequence 70, Application US/09502967
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth

```

```

1  APPLICANT: MCCOY, John M.
2  APPLICANT: Lavallie, Edward R.
3  APPLICANT: Collins-Racle, Lisa A.
4  APPLICANT: Evans, Cheryl
5  APPLICANT: Merberg, David
6  APPLICANT: Treacy, Maurice
7  APPLICANT: Spaulding, Vilki
8  APPLICANT: Genetics Institute, Inc.
9  TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
10 FILE REFERENCE: GI 6021-30X
11 CURRENT APPLICATION NUMBER: US/09/502,967
12 CURRENT FILING DATE: 2000-02-11
13 NUMBER OF SEQ ID NOS: 306
14 SOFTWARE: PatentIn Ver. 2.0
15 SEQ ID NO 70
16 LENGTH: 615
17 TYPE: PRT
18 ORGANISM: Homo sapiens
19 US-09-502-967-70

```


QY	406	AGGCTTTAAATCCAGTGTGACACAGTTGAGGAAAGAGCTGGAGAGCGAGGGGAGCGCACTT	4655
Db	41	LYSLALALEUILLGGINCYASSPGIDLEUATRYGSLUENUGLUAAGLUALAGLUATG	60
QY	466	GAATAAGAACTTGCACTTCACAGCAAGAGAAAAGGGCCATTGGAAGAAGCATGATGAAAAAG	5255
Db	61	GIUTYSGIULEUADLASEGIDNGINLUYARGLAILEGIULUYASPMETUETLYLS	80
QY	526	GAATAACGAAAAGAAAGGAGTACATGGGATGCCAAAGATGTTGATCTTGTCTCAGATATTT	5855
Db	81	GIUIETHTLYSGIUAUGIUTRYMETDIYSEUTLYSEUTLEUILEUSERGILNAINLLE	1005
QY	586	GCCCAACTGAGAGGCCACAGTGGAAAAGGTTTACAAAAGGAAAGATTTCCAGCTATTATTCAA	6455
Db	101	ALAGINLEUGLUALAGLINALGILUYVALHTRHYSGLUYSLIESERLAIIAIEASINGLN	1205
QY	646	CTGAGAGAAATTCANAAAGCCAGCTGGCTTTCTCGGGAATAATGCATGTCCAAAAGTGTGTGA	7055
Db	121	LEUGIUGIUIILEGINSEGLIDLEALASERGILUETAPVALTHRLYVALCYSGILY	1405
QY	706	GAATGGCCCTATTCAGCTGCAATATTAACCAACATCGAGAAAGATGAGGCAAAAAGACAC	7655
Db	141	GIUMETATGYTGILUENALASHYSTRHYSINMETGIUYASPGIUALAGIULUYSGILNIS	1605
QY	766	AGAGAGTTTCAGAGCAAAACATACAGAGGATTTGAAAATTAAAGATTCAGAAATAGAGAAA	8255
Db	161	ARGIUIPHREARGIATALYETHTASHNARGSP**GIUIETLYSASPGILNUIILEGIULYS	1805
QY	826	TTGAAATATAGAACTGGATGGAAGCAACACACTTTGGAAACGAGGACGACAGAGGACGCC	8855
Db	181	LEUATGILIEGLIUENAUSPGILUSERLYSGILNHSLEUGIUGIUGIUGIUGIULYSLAIALA	2005
QY	886	CTGGCCAGA 894	
Db	201	LEUALIATRG 203	
RESULT 6			
US-60-389-987-766			
: Sequence 766, Application US/60389987			
: GENERAL INFORMATION:			
: APPLICANT: Ghosh, Soumitra S.			
: APPLICANT: Fahy, Eoin D.			
: APPLICANT: Zhang, Bing			
: APPLICANT: Gibson, Bradford W.			
: APPLICANT: Taylor, Steven W.			
: APPLICANT: Glenn, Gary M.			
: APPLICANT: Warnock, Dale E.			
: TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION			
: FILE REFERENCE: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME			
: CURRENT APPLICATION NUMBER: US/60/389,987			
: CURRENT FILING DATE: 2002-06-17			
: NUMBER OF SEQ ID NOS: 3025			
: SOFTWARE: FastSeq for Windows Version 4.0			
: SEQ ID NO 766			
: LENGTH: 393			
: TYPE: PRT			
: ORGANISM: Homo sapiens			
US-60-389-987-766			
Alignment Scores:			
Pred. NO.: 5.27e-53			
Score: 841.00			
Percent Similarity: 100.00%			
Best Local Similarity: 100.00%			
Query Match: 30.78%			
DB: 27			
Gaps: 0			
US-09-502-945-1 (1-1552) x US-60-389-987-766 (1-393)			
QY	16	GAGAGGCTTAATACTTACTTATGAGGAAAAGTGTAATTTAGAGAAATCCCAATTCAGATT	75

Db	226	gltulysleuylsleuthrttygluglulubyscsglulieglugluserglnleuyls	phe	2455
Qy	76	TTGAGAACGACTTAACTGTAATATFCAGAGAACTTGTAGAGATCTTTAAAGACAACTTAAG	1355	
Db	246	leuatargasnspdleualaglutyrglnarphthrcysglnaspleuylsglulglnleuyls	2655	
Qy	136	CTTAAAGAAATTTCTTCTGCGGCGCTAAATCTTTGAACCGTGTGGTGCTTTGTTGAAA	1955	
Db	266	hlslysgluphleleuialeualalsnthrlyssaslnayvalglyglyleucysleuyls	2855	
Qy	196	TGTGCTCAGCATGAGACTGTTCTTTCCCAAAACCCATCTAATCTTCAATATGACAGACCATC	2555	
Db	286	cysalialaglnhlsiglualalavalleuserclnrhrhsthrasnvahlsmetglnthrlle	3055	
Qy	256	GAAAGACTGGTTTAAAGAAAGAGATGACTTGATGTCTGCACCTAGTTTCCGTAAGAGCAGC	3155	
Db	306	gluarlgleuvallysgluarlgaspaspseumetseralaleuvalservalarqserSer	3255	
Qy	316	TTGGCAGATATCCGACCAAGAAGAGCAGTCGCTTATGACACGCGTAACAAGTTTGCAA	3755	
Db	326	leualiaasrpthnglnlarlglnalaserlatryrglulglnvallysglnvalleuGln	3455	
Qy	376	ATATCTGAGGAAGCCAAATTTTGAAAAAACAAGCCTTTAATCCAGTGTGACAGTTGAGG	4355	
Db	346	lleserlglnlualiasnphelulysrthryslalaleuileglnycysaprglnleuarg	3655	
Qy	436	AAGGACTGTGAGAGCGCAGCGGCGGCGACTTGAATAAAGAACTTGATCTCAGCAAGAGAAA	4955	
Db	366	lysrgluleuGlnuarlglnalagluarlglenGlnlysgluleualaserGlnGlnGlnlys	3855	
Qy	496	AGGGCCATTTGAGAAAGACATGATG 519		
Db	386	ArghAlaIleGlnlyspasmetchet 393		
RESULT 7				
US-60-412-418-766				
Sequence 766, Application US/60412418				
GENERAL INFORMATION:				
APPLICANT: Ghosh, Soumitra S.				
APPLICANT: Fahy, Eoin D.				
APPLICANT: Zhang, Bing				
APPLICANT: Gibson, Bradford W.				
APPLICANT: Taylor, Steven W.				
APPLICANT: Glenn, Gary M.				
APPLICANT: Warnock, Dale E.				
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION				
TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME				
FILE REFERENCE: 660088.465P3				
CURRENT APPLICATION NUMBER: US/60/412,418				
CURRENT FILING DATE: 2002-09-20				
NUMBER OF SEQ ID NOS: 3025				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 766				
LENGTH: 393				
TYPE: PRT				
ORGANISM: Homo sapiens				
US-60-412-418-766				
Alignment Scores:				
Pred. No.:		5,27e-53	Length:	393
Score:		841.00	Matches:	168
Percent Similarity:		100.00%	Conservative:	0
Best Local Similarity:		100.00%	Mismatches:	0
Query Match:		30.78%	Indels:	0
DB:		27	Gaps:	0
US-09-502-945-1 (1-1552) x US-60-412-418-766 (1-393)				
Qy	16	GAGAGCTAAACTTACTTATGAGGAAAGGTGAAGTTGAGAGATCCCAATTGAAGTTT	75	
Db	226	gltulysleuylsleuthrttygluglulubyscsglulieglugluserglnleuyls	2455	
Qy	76	TTGAGAACGACTTAACTGTAATATFCAGAGAACTTGTAGAGATCTTTAAAGACAACTTAAG	1355	
Db	246	leuatargasnspdleualaglutyrglnarphthrcysglnaspleuylsglulglnleuyls	2655	
Qy	136	CTTAAAGAAATTTCTTCTGCGGCGCTAAATCTTTGAACCGTGTGGTGCTTTGTTGAAA	1955	
Db	266	hlslysgluphleleuialeualalsnthrlyssaslnayvalglyglyleucysleuyls	2855	
Qy	196	TGTGCTCAGCATGAGACTGTTCTTTCCCAAAACCCATCTAATCTTCAATATGACAGACCATC	2555	
Db	286	cysalialaglnhlsiglualalavalleuserclnrhrhsthrasnvahlsmetglnthrlle	3055	
Qy	256	GAAAGACTGGTTTAAAGAAAGAGATGACTTGATGTCTGCACCTAGTTTCCGTAAGAGCAGC	3155	
Db	306	gluarlgleuvallysgluarlgaspaspseumetseralaleuvalservalarqserSer	3255	
Qy	316	TTGGCAGATATCCGACCAAGAAGAGCAGTCGCTTATGACACGCGTAACAAGTTTGCAA	3755	
Db	326	leualiaasrpthnglnlarlglnalaserlatryrglulglnvallysglnvalleuGln	3455	
Qy	376	ATATCTGAGGAAGCCAAATTTTGAAAAAACAAGCCTTTAATCCAGTGTGACAGTTGAGG	4355	
Db	346	lleserlglnlualiasnphelulysrthryslalaleuileglnycysaprglnleuarg	3655	
Qy	436	AAGGACTGTGAGAGCGCAGCGGCGGCGACTTGAATAAAGAACTTGATCTCAGCAAGAGAAA	4955	
Db	366	lysrgluleuGlnuarlglnalagluarlglenGlnlysgluleualaserGlnGlnGlnlys	3855	
Qy	496	AGGGCCATTTGAGAAAGACATGATG 519		
Db	386	ArghAlaIleGlnlyspasmetchet 393		

|||||
Db 246 LeuArGaNaSpLeuAlaGluTyrGlnArgThrCysGluAspLeuTyrGlnLeuLys 265
OY 136 CATAAAGAAATTTCTTGGCTGCTAATACTGTAAACCGTGTGGCTTTGTTTGA 195
Db 266 HisTyrGlnPheLeuLeuAlaAlaSerThrCysAsnArgValGlyGlyLeuCysLeuLys 285
OY 196 TGTGCTCAGCTGAAGCTGTCTTCCCAACCCATCAATGTGTCATATGTCAGAGCCATC 255
Db 286 CysAlaGlnHisGluAlaValLeuSerGlnThrHisThrAsnValHisMetGlnThrIle 305
OY 256 GAAAGACTGTGTTAAAGAAAGAGATGACTGTGTCGTGCTGCTAGTTCCTGTAAGAGCAGC 315
Db 306 GluArgLeuValLysGluArgAspAspLeuMetSerAlaLeuValSerValArgSer 325
OY 316 TTGGCAGATACCGCAGCAAGAACCAAGTCTTATGACAGGGTGAACAAGTTTGC 375
Db 326 LeuAlaAspThrGlnGlnArgGlnValAserAlaTyrGlnGlnValLysGlnValLeuGln 345
OY 376 ATATCTGAGGAAGCCATTTTGAAAAACCAAGCTTTAAATCCAGTGTGACAGTGTGAG 435
Db 346 IleSerGlnGluAlaAsnPheGluTyrThrLysAlaLeuIleGlnCysAspLeuLeuArg 365
OY 436 AAGGAGCTGAGAGGCGAGCGAGCGACTTGAAAAAGAACTTGCTCAGCAAGAGAAA 495
Db 366 LysGluLeuGluArgGlnAlaGlnArgLeuGluLysGluLeuAlaSerGlnGlnLys 385
OY 496 AGGCCATTGAGAAAACATGATG 519
Db 386 ArgAlaIleGlnLysAspMetMet 393

RESULT 8

PCT-US01-03800A-1921
Sequence 1921, Application PC/TUS0103800A
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
APPLICANT: Tang, Y, Tom et al
TITLE OF INVENTION: Novel Nucleic Acids and Polypeptides
FILE REFERENCE: 21272-030
CURRENT APPLICATION NUMBER: PCT/US01/03800A
CURRENT FILING DATE: 2001-02-05
PRIORITY APPLICATION NUMBER: 09/560,875
PRIORITY FILING DATE: 2000-04-27
PRIORITY APPLICATION NUMBER: 09/496,914
PRIORITY FILING DATE: 2000-02-03
NUMBER OF SEQ ID NOS: 2700
SOFTWARE: Custom
SEQ ID NO 1921
LENGTH: 172
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-03800A-1921

Alignment Scores:

Pred. No.: 1.02e-30 Length: 172
Score: 537.00 Matches: 117
Percent Similarity: 70.21% Conservative: 15
Best Local Similarity: 62.23% Mismatches: 40
Query Match: 19.66% Indels: 16
DB: 1 Gaps: 3

US-09-502-945-1 (1-1552) x PCT-US01-03800A-1921 (1-172)

OY 670 GCTTCTGGGGAATGATGTCAACAAGGTGTGTGGAAGAAATGCCGTATCAGCTGAATAA 729
Db 1 AlaserArgGluMetAspValThrLysValLysGlyGluMetArgTyrGlnLeuAsnLys 20
OY 730 ACCAATGAGAGAGAGAGAGCAAGAAAGAGCAGACAGAGTTCAGAGCAAAAACATAC 789
Db 21 ThrSnmctGluLysAspGluAlaGluLysGlnHisArgGlnPheArgAlaLysThrAsn 40
OY 790 AGGATCTTGAATTAAGATCAGAGAAATAGAGAAATTGAGAAATAGAACTGGATGAAGC 849
|||||

Db 41 ArgAspLeuGluLysLysAspGlnGluLysLeuArgIleGluLeuAspLysSer 60
OY 850 AAACACACTTGGAGACAGAGACAGAGAGAGCCCTGCCAGAGAGAGATGCTGTGA 909
Db 61 LysGlnHisLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 80
OY 910 CTAACAGAACTGTGGGGAATCTGAGCACCACCACTGCACCTCCACAGATGTAATACCT 969
Db 81 LeuThrGluLeuLeuGlyGlnSerGlnHisGlnLeuHisLeuThrArgGlnGluLysAsp 100
OY 970 CAATCTAGTCAGAGAAAAAGATATACATATGATTAATTTGGAAAGTTCAGAGAAAT 1029
Db 101 SerIleGlnGln-----SerPheSerLysGluAlaLysAlaGlnAlaLeuGln 116
OY 1030 GAAGATTTGAGAGAACACTGTGTCCAGCATGGAGAGATGACATGACATCAAGCAAG 1089
Db 117 AlaGlnGlnArgGlnGln-----GluLeuThrGlnLys 127
OY 1090 CTAAGGAGCTGGATTAACACAGCAGCCAGCCAGCCAGCTGTGCTCCTCAGC 1149
Db 128 IleGlnGlnMetGluAlaGlnHisAspLysThrGlnAsnGlnGlnTyrLeuLeuLeuThr 147
OY 1150 AAGCAGAACACAGCTTCTCTGAGAGAGCAGCCCTGTGGAAGAGTGGACCGCTGCG 1209
Db 148 SerGlnAsnThrPheLeuThrLys-----LeuLysGlnGluCysCysThrLeuAla 164
OY 1210 ACCCAGTTACCCAGCATCCACAA 1233
Db 165 LysLysLeuGlnGlnIleSerGln 172

RESULT 9

US-09-834-366-14401
Sequence 14401, Application US/09834366
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephanie
APPLICANT: Tanaka, Hiroaki
APPLICANT: Dumas Milne Edwards, Jean Baptiste
APPLICANT: Jobert, Severin
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: 81.052.REG
CURRENT APPLICATION NUMBER: US/09/834,366
CURRENT FILING DATE: 2001-04-13
PRIORITY APPLICATION NUMBER: US 60/197,873
PRIORITY FILING DATE: 2000-04-18
NUMBER OF SEQ ID NOS: 52153
SOFTWARE: Patent.pm
SEQ ID NO 14401
LENGTH: 138
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: 126
OTHER INFORMATION: Xaa - Asp, Glu
US-09-834-366-14401

Alignment Scores:

Pred. No.: 1.74e-24 Length: 138
Score: 452.00 Matches: 96
Percent Similarity: 74.50% Conservative: 15
Best Local Similarity: 64.43% Mismatches: 24
Query Match: 16.54% Indels: 14
DB: 22 Gaps: 3

US-09-502-945-1 (1-1552) x US-09-834-366-14401 (1-138)

OY 709 ATGGGCTATCGCTGAATTAATAACCAATGAGAGAGATGAGAGAGAGAGAGAGAGAG 768
Db 1 MetArgTyrGlnLeuAsnLysThrAsnMetGluLysAspGluAlaGluLysGlnHisArg 20
OY 769 GAGTTCAGAGCAAAACATCAAGGAGATCTGAATTAAGATCAGGAATAGAGAAATTG 828
|||||

```
Db 21 GluPheArGAlaLysThrAsnArgSpLeuGluLLeLysAspInLuiLleLysLeu 40
QY 829 AGAATGAACTGGATGAAACCAAACTTGTGACAGAGCAGAGCAAGCCCTG 888
    |||||
Db 41 ArgILeLLeuAspLysSerLysGlnHisLeuGlnLuiLleGlnLysAlaLLeu 60
QY 889 GCCAGAGAGAGTGCCTGACACTAACAGAACTCTGGCGAATCTGAGCACAATGCGAC 948
    |||||
Db 61 AlaArgLuiLLeuLysLeuArgLeuThrGlnLeuLeuGlnLuiLleSerGlnHisGlnLeuHis 80
QY 949 CTCACAGACTGCAATAGCTCAACTCAGTCAAGAAAGATATCATGTGATTAATG 1008
    |||||
Db 81 LeuThrArgLInLuiLysAspSerLLeGlnGln-----SerPheSerLysGlu 96
QY 1009 GGAAGCTTACAGAGAAATGAAGATTGGAGAACACTGTGTCCAGCATGGAGAGTA 1068
    |||||
Db 97 AlaLysAlaGlnLLeuGlnAlaGlnGlnArgGlnLui----- 109
QY 1069 CATGACAGCATGAACCAAGGCTTAAGCAGCTGGAT---AAGCACAAGCCAGCCACAGCC 1125
    |||||
Db 110 -----GluLeuThrGlnLysLLeGlnGlnMetGlnLuiAlaGlnHisAspLysThr**Asn 127
QY 1126 CACGAGCTGTGCAGCTCCACAGCAAG 1152
Db 128 GluGlnTyrlLeuLeuLeuThrSerArg 136

RESULT 10
US-60-197-873-14401
; Sequence 14401, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Benjamin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Malne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81 US1.PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 14401
; LENGTH: 138
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 126
; OTHER INFORMATION: Xaa = Asp, Glu
US-60-197-873-14401

Alignment Scores:
Pred. No.: 1.74e-24 Length: 138
Score: 452.00 Matches: 96
Percent Similarity: 74.508 Conservative: 15
Best Local Similarity: 64.438 Mismatches: 24
Query Match: 16.54% Indels: 14
DB: 27 Gaps: 3

US-09-502-945-1 (1-1552) x US-60-197-873-14401 (1-138)
QY 709 ATGCGCTATCAGCTGAATTAACCAACATGAGAGAGTGGCAGAAAAAGAGCAGACA 768
    |||||
Db 1 MetArgTyrlGlnLeuAsnLysThrAsnMetGlnLysAspLuiLleGlnLysGlnHisArg 20
QY 769 GAGTTGAGAGCAAAACTAAGAGGATCTTGAATTAAGATACAGAAATAGGAATTC 828
    |||||
Db 21 GluPheArgAlaLysThrAsnArgSpLeuGlnLLeLysAspInLuiLleGlnLysLeu 40
QY 829 AGAATGAACTGGATGAAACCAAACTTGTGACAGAGCAGAGCAAGCCCTG 888
    |||||
Db 41 ArgILeLLeuAspLysSerLysGlnHisLeuGlnLuiLleGlnLysAlaLLeu 60
```

```
QY 889 GCCAGAGAGAGTGCCTGACACTAACAGAACTCTGGCGAATCTGAGCACAATGCGAC 948
    |||||
Db 61 AlaArgLuiLLeuLysLeuArgLeuThrGlnLeuLeuGlnLuiLleSerGlnHisGlnLeuHis 80
QY 949 CTCACAGACTGCAATAGCTCAACTCAGTCAAGAAAGATATCATGTGATTAATG 1008
    |||||
Db 81 LeuThrArgLInLuiLysAspSerLLeGlnGln-----SerPheSerLysGlu 96
QY 1009 GGAAGCTTACAGAGAAATGAAGATTGGAGAACACTGTGTCCAGCATGGAGAGTA 1068
    |||||
Db 97 AlaLysAlaGlnLLeuGlnAlaGlnGlnArgGlnLui----- 109
QY 1069 CATGACAGCATGAACCAAGGCTTAAGCAGCTGGAT---AAGCACAAGCCAGCCACAGCC 1125
    |||||
Db 110 -----GluLeuThrGlnLysLLeGlnGlnMetGlnLuiAlaGlnHisAspLysThr**Asn 127
QY 1126 CACGAGCTGTGCAGCTCCACAGCAAG 1152
Db 128 GluGlnTyrlLeuLeuLeuThrSerArg 136

RESULT 11
PCT-US02-09921-968
; Sequence 968, Application PC/TUS0209921
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: DAFEO, Abel
; APPLICANT: JONES, Anissa L.
; APPLICANT: TRAN, Alanna-Phung B.
; APPLICANT: DAHL, Christopher R.
; APPLICANT: GIERZEN, Darryl
; APPLICANT: CHINN, Joyce
; APPLICANT: DUFOUR, Gerard E.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: TUASON, Olivia
; APPLICANT: YAP, Pierre E.
; APPLICANT: AMSEY, Stefan R.
; APPLICANT: DAUGHERTY, Sean C.
; APPLICANT: DAM, Tam C.
; APPLICANT: LIU, Tommy F.
; APPLICANT: NGUYEN, Duy-Viet An
; APPLICANT: KLEBEID, Yael
; APPLICANT: KERSTIN JR., Edward H.
; APPLICANT: PERALTA, Careyna H.
; APPLICANT: DAVID, Marie H.
; APPLICANT: LEWIS, Samantha A.
; APPLICANT: CHEN, Alice J.
; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1232 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09921
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
60/291,829; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
; NUMBER OF SEQ ID NOS: 1146
; SOFTWARE: PERL Program
; SEQ ID NO 968
; LENGTH: 94
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No: LT:474414.28.off1:2001MAY17
PCT-US02-09921-968
```

Alignment Scores:

Pred. No.: 2,71e-24 Length: 94
 Score: 449.00 Matches: 81
 Percent Similarity: 93.10% Conservatve: 0
 Best Local Similarity: 93.10% Mismatches: 0
 Query Match: 16.30% Indels: 6
 DB: 1 Gaps: 1

US-09-502-945-1 (1-1552) x PCT-US02-09921-968 (1-94)

OY 1223 CTGGGTAAC-----TGGTCGCGACCGGTCACCTTTCCAGACG 1182
 DB 8 LeuGIaAnIIleValArgProHisLeuTyrValArgSerArgSerThrSerSerAspArg 27
 OY 1181 CTCTGSCCTTCAGAGAGAGCGCTTCCTGCTGAGAGAGTGCAGACGCTCGGGCT 1122
 DB 28 LeuCySLeuSerArgArgSerTyrPheCySLeuLeuArgSerCySThrSerCyStrpAla 47
 OY 1121 GTGGCGCTGGCTGTCTTTCACAGCTGCCTTAGCCTTTGCTTCATGCTCATGTACTCTC 1062
 DB 48 ValAlaIrrpleuCySLeuSerSerCySLeuSerLeuCySPhelIleValSerCyStrIreu 67
 OY 1061 CCAGCTGACACACTGTCTCCAAATTCCTCATTTCTCTCTGTAACCTTCCAAATTTA 1002
 DB 68 ProCyStrPThrHisCySserSerSerSerPheLeuLeuCySAsnPhroAsnleu 87
 OY 1001 TCATATGTATACCTTTTCT 981
 DB 88 SerTyValTyLeuPheSer 94

RESULT 12

US-09-834-366-21271
 : Sequence 21271, Application US/09834366
 : GENERAL INFORMATION:
 : APPLICANT: Benjamin, Stephane
 : APPLICANT: Tanaka, Hiroaki
 : APPLICANT: Dumas Milne Edwards, Jean Baptiste
 : APPLICANT: Jobert, Severin
 : APPLICANT: Giordano, Jean-Yves
 : TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 : FILE REFERENCE: 81.US2.REG
 : CURRENT APPLICATION NUMBER: US/09/834,366
 : CURRENT FILING DATE: 2001-04-13
 : PRIOR APPLICATION NUMBER: US 60/197,873
 : PRIOR FILING DATE: 2000-04-18
 : NUMBER OF SEQ ID NOS: 52153
 : SOFTWARE: Patent.pm
 : SEQ ID NO 21271
 : LENGTH: 61
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-09-834-366-21271

Alignment Scores:

Pred. No.: 4.16e-14 Length: 61
 Score: 310.00 Matches: 61
 Percent Similarity: 98.39% Conservatve: 0
 Best Local Similarity: 98.39% Mismatches: 0
 Query Match: 11.35% Indels: 1
 DB: 22 Gaps: 0

US-09-502-945-1 (1-1552) x US-09-834-366-21271 (1-61)

OY 1078 ATGAAGCAAGGCTAAGCAGCTGATTAAGCAGACAGCCAGCCAGCCAGCTGGTG 1137
 DB 1 MetLysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20
 OY 1138 CAGCTCTCAGACAGACAGACCTTCTCTGAGAGAGAGAGCTTCGGAGAGGTTG 1197
 DB 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuLeuGlnArgInser-CysArgLysArgTr 40
 OY 1198 GACCGGCTGCGGACCACTAACCCAGCATGCCAATCTGATGTGACTGATGTAAC 1257
 DB 60 rGlu 61

DB 40 pThrGlyCySglyProSerTyProAlaCySHisAsnLeuIleAlaAspLeuAspGlyTh 60

OY 1258 AGAG 1261

DB 60 rGlu 61

RESULT 13

US-60-197-873-21271
 : Sequence 21271, Application US/60197873
 : GENERAL INFORMATION:
 : APPLICANT: Benjamin, Stephane
 : APPLICANT: Tanaka, Hiroaki
 : APPLICANT: Dumas Milne Edwards, Jean Baptiste
 : APPLICANT: Jobert, Severin
 : APPLICANT: Giordano, Jean-Yves
 : TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 : FILE REFERENCE: 81.US1.PRO
 : CURRENT APPLICATION NUMBER: US/60/197,873
 : CURRENT FILING DATE: 2000-04-18
 : NUMBER OF SEQ ID NOS: 52153
 : SOFTWARE: Patent.pm
 : SEQ ID NO 21271
 : LENGTH: 61
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-60-197-873-21271

Alignment Scores:

Pred. No.: 4.16e-14 Length: 61
 Score: 310.00 Matches: 61
 Percent Similarity: 98.39% Conservatve: 0
 Best Local Similarity: 98.39% Mismatches: 0
 Query Match: 11.35% Indels: 1
 DB: 27 Gaps: 0

US-09-502-945-1 (1-1552) x US-60-197-873-21271 (1-61)

OY 1078 ATGAAGCAAGGCTAAGCAGCTGATTAAGCAGACAGCCAGCCAGCCAGCTGGTG 1137
 DB 1 MetLysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20
 OY 1138 CAGCTCTCAGACAGACAGACCTTCTCTGAGAGAGAGAGCTTCGGAGAGGTTG 1197
 DB 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuLeuGlnArgInser-CysArgLysArgTr 40
 OY 1198 GACCGGCTGCGGACCACTTACCCAGCATGCCAATCTGATGTGACTGATGTAAC 1257
 DB 40 pThrGlyCySglyProSerTyProAlaCySHisAsnLeuIleAlaAspLeuAspGlyTh 60
 OY 1258 AGAG 1261
 DB 60 rGlu 61

RESULT 14

US-09-621-976-5141
 : Sequence 5141, Application US/09621976
 : GENERAL INFORMATION:
 : APPLICANT: Dumas Milne Edwards, J.B.
 : APPLICANT: Jobert, S.
 : APPLICANT: Giordano, J.Y.
 : TITLE OF INVENTION: ESTs and Encoded Human Proteins.
 : FILE REFERENCE: GENSET.054PR2
 : CURRENT APPLICATION NUMBER: US/09/621,976
 : CURRENT FILING DATE: 2000-07-21
 : NUMBER OF SEQ ID NOS: 19335
 : SOFTWARE: Patent.pm
 : SEQ ID NO 5141
 : LENGTH: 62
 : TYPE: PRT
 : ORGANISM: Homo sapiens
 US-09-621-976-5141

Alignment Scores:

Pred. No.: 2.26e-13 Length: 62
Score: 300.00 Matches: 59
Percent Similarity: 98.33% Conservative: 0
Best Local Similarity: 98.33% Mismatches: 0
Query Match: 10.98% Indels: 1
DB: 20 Gaps: 0

US-09-502-945-1 (1-1552) x US-09-621-976-5141 (1-62)

OY 1078 ATGAGCAAGGCTAAGCAGCTGATTAAGCAGCCAGCCAGCCAGCAGCTGCTG 1137
|||||
Db 1 MetLysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20

OY 1138 CAGCTCTCAGCAAGCAGAACCAAGCTTCTCCTGAGAGCCAGCCTGTGGAAGGTG 1197
|||||
Db 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuGlnArgGlnSer-CysArgLysArgTr 40

OY 1198 GACCGGCTGCGGACCCAGTTTACCAGCATGCCCAATCTGATTGCTGACCTGGATGGA 1255
|||||
Db 40 pThrGlyCysGlyProSerTyrProAlaCysHisAsnLeuIleAlaAspLeuAspGly 59

RESULT 15

US-60-147-499-5141

; Sequence 5141, Application US/60147499
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/60/147,499
; CURRENT FILING DATE: 1999-08-05
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 5141
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-147-499-5141

Alignment Scores:

Pred. No.:	2.26e-13	Length:	62
Score:	300.00	Matches:	59
Percent Similarity:	98.33%	Conservative:	0
Best Local Similarity:	98.33%	Mismatches:	0
Query Match:	10.98%	Indels:	1
DB:	27	Gaps:	0

US-09-502-945-1 (1-1552) x US-60-147-499-5141 (1-62)

OY 1078 ATGAGCAAGGCTAAGCAGCTGATTAAGCAGCCAGCCAGCCAGCAGCTGCTG 1137
|||||
Db 1 MetLysGlnArgLeuArgGlnLeuAspLysHisSerGlnAlaThrAlaGlnGlnLeuVal 20

OY 1138 CAGCTCTCAGCAAGCAGAACCAAGCTTCTCCTGAGAGCCAGCCTGTGGAAGGTG 1197
|||||
Db 21 GlnLeuLeuSerLysGlnAsnGlnLeuLeuGlnArgGlnSer-CysArgLysArgTr 40

OY 1198 GACCGGCTGCGGACCCAGTTTACCAGCATGCCCAATCTGATTGCTGACCTGGATGGA 1255
|||||
Db 40 pThrGlyCysGlyProSerTyrProAlaCysHisAsnLeuIleAlaAspLeuAspGly 59

Search completed: March 21, 2003, 13:42:51
Job time : 174.803 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: March 21, 2003, 13:00:02 ; Search time 16.6182 Seconds
(without alignments)
9985.861 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732
Sequence: 1 ctcttgatgacatccgagaa.....aaatgacttttaagaaga 1552

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 442306

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+n2p.model -DEV=xlp
-O/cgnt2_1/USOTO.spool/US0502945/runtat_14032003_101101_19301/app_query.fasta.1.10979
-DB=published.Applications_AA -QEMT=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62
-TRANS=human40.csl -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US0502945.0CGN.1.1.60.0.runtat_14032003_101101_19301
-NCPU=6 -ICPU=3 -NO_XLPHY -NO_MMAR -LARGOQUERY -NEG_SCORES=0 -WAT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:*

1: /cgnt2_6/prodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgnt2_6/prodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgnt2_6/prodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgnt2_6/prodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgnt2_6/prodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgnt2_6/prodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgnt2_6/prodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgnt2_6/prodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgnt2_6/prodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgnt2_6/prodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgnt2_6/prodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgnt2_6/prodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgnt2_6/prodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgnt2_6/prodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242.5	8.9	2139	9	US-10-023-219-4
2	242.5	8.9	2139	10	US-09-727-384-6
3	238.5	8.7	868	9	US-09-884-001-19
4	231.5	8.5	2125	10	US-09-919-172-29

5	229.5	8.4	909	9	US-09-925-299-988	Sequence 988, App
6	229.5	8.4	909	10	US-09-925-299-988	Sequence 988, App
7	229	8.4	2354	9	US-09-820-843A-113	Sequence 113, App
8	228	8.3	2310	9	US-09-991-496-120	Sequence 120, App
9	228	8.3	2310	9	US-09-820-843A-114	Sequence 114, App
10	228	8.3	2310	10	US-09-874-923-120	Sequence 120, App
11	226.5	8.3	888	9	US-09-893-519A-73	Sequence 73, App
12	224.5	8.2	1805	9	US-09-820-843A-73	Sequence 73, App
13	222.5	8.1	1045	10	US-09-815-242-10617	Sequence 10617, A
14	220	8.1	2053	9	US-10-017-216-2	Sequence 2, App
15	218.5	8.0	660	10	US-09-864-761-47959	Sequence 47959, A
16	218.5	8.0	1945	9	US-09-927-597-2	Sequence 2, App
17	218.5	8.0	1597	9	US-09-927-597-4	Sequence 4, App
18	215.5	7.9	1597	9	US-10-017-216-6	Sequence 6, App
19	215.5	7.9	2055	9	US-10-017-216-4	Sequence 4, App
20	215	7.9	1641	9	US-10-017-216-5	Sequence 5, App
21	214.5	7.9	1958	12	US-10-028-946-4	Sequence 4, App
22	214.5	7.9	2054	12	US-10-028-946-2	Sequence 2, App
23	213	7.8	900	12	US-10-071-751-21	Sequence 21, App
24	208	7.6	443	1	US-08-325-278-6	Sequence 6, App
25	207.5	7.6	691	9	US-10-028-972-16	Sequence 16, App
26	207.5	7.6	691	9	US-10-121-049-16	Sequence 16, App
27	207.5	7.6	691	9	US-10-123-904-16	Sequence 16, App
28	207.5	7.6	691	9	US-10-140-470-16	Sequence 16, App
29	207.5	7.6	691	9	US-10-175-746-16	Sequence 16, App
30	207.5	7.6	691	9	US-10-176-918-16	Sequence 16, App
31	207.5	7.6	691	9	US-10-176-921-16	Sequence 16, App
32	207.5	7.6	691	9	US-10-137-865-16	Sequence 16, App
33	207.5	7.6	691	9	US-10-140-474-16	Sequence 16, App
34	207.5	7.6	691	9	US-10-142-431-16	Sequence 16, App
35	207.5	7.6	691	9	US-10-143-114-16	Sequence 16, App
36	207.5	7.6	691	9	US-10-140-002-16	Sequence 16, App
37	207.5	7.6	691	9	US-10-142-419-16	Sequence 16, App
38	207.5	7.6	691	9	US-10-123-262-16	Sequence 16, App
39	207.5	7.6	691	9	US-10-142-423-16	Sequence 16, App
40	207	7.6	830	9	US-10-033-245-7	Sequence 7, App
41	207	7.6	830	9	US-10-033-223-7	Sequence 7, App
42	207	7.6	830	9	US-10-033-167-7	Sequence 7, App
43	207	7.6	830	9	US-10-033-244-7	Sequence 7, App
44	207	7.6	830	9	US-10-033-435-7	Sequence 7, App
45	207	7.6	830	9	US-10-032-990-7	Sequence 7, App

ALIGNMENTS

RESULT 1
US-10-023-219-4
; Sequence 4, Application US/10023219
; Publication No. US20030032592A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Cimbro, Daniel M.
; APPLICANT: Heichman, Karen
; APPLICANT: Bartel, Paul L.
; TITLE OF INVENTION: Protein-Protein Interactions
; FILE REFERENCE: 2318-282-II
; CURRENT APPLICATION NUMBER: US/10/023, 219
; PRIOR APPLICATION NUMBER: 2001-12-20
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 2139
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-023-219-4

Alignment Scores:
Pred. No.: 3 21e-10
Score: 242.50
Percent Similarity: 41.86%
Best Local Similarity: 24.21%
Length: 2139
Matches: 107
Conservative: 78
Mismatch: 158

Query Match:	8.88%	Indels:	99
DB:	9	Gaps:	18
US-09-502-945-1 (1-1552) x US-10-023-219-4 (1-2139)			
QY 10 GCATCCGGAAGCTAAACCTTACTTATGAGGAAAAGTGTGAAATTTGAGAGATCCCAATTCG			69
Db 656 ALaGInGInaSMetLySGInaRhISGInaSnInGInaRhGrThLeuGInGInaIle 67575			
QY 70 AAGTTTGTGAGAAAGCACTTAAGTATACAGAACTTGTGAGATCTTAAAGCAAA 1299			
Db 676 SerAspLeuLySaSnGInUlealIGInUleSnInGInUlealalValaLeuLySGIn 694			
QY 130 CTAAGCACTAAAGAAATTCTTCGGCGCATATCTGTGAACCGTGTGGTGGCTTTGT 1899			
Db 694 ----- 694			
QY 190 TTGAAATGCTGCAGCATGAGCTGTCTTCCCAACCCATACATAATGTTCAATGAC 249			
Db 695 -----AlaInSHISGInaIaIThCysaRhGInISGInGInUlySGInLeuGIn 71111			
QY 250 AACATCGAAAGACTGGTTAAAGAAAGATGACTGTATGTCGCATA----- 297			
Db 712 Val-----LysLeuGInGInUlyGInUlyThRhISLeuGInGInUlySerAtrGleuGInIS 7299			
QY 298 ---GTTTCCGTAAGAGAGCTGTGGCAGATATACGAGCAAAAGAGAAAGTGGCTTAAGAA 354			
Db 730 GlUmetGInUleLySaIaATrGInUleThRhInaIaGIn-----AlaSerPheGIn 745			
QY 355 CAGGTGAAACAAGTTTGTTCGATATATTCGAGGAACCCAAATTTTGAAGAAAACCAAGGCTTTA 4144			
Db 746 ArgGInaUatrgGInUlyLeuGInSerSerAlaTrPheGIn---GInUlySaIaATrGInUleu 764			
QY 415 ATCCAGTGTGACCAAGTTGAGAGAGAGCTGGAGGAGCGGCGAGCACTTGAAGAA 4744			
Db 765 ThrGIn-----GInUleuGIn-----GInPheInSGInGIn 775			
QY 475 CTTCGATCTCGAACAGAAAGGCGCATTCGAAACACATGATGAAAGAAATACG 534			
Db 776 LeuThrSerLeuValGInUlySHISIThLeuGInUlySGInUleuAtrGInUlySerLeu 7959			
QY 535 AAAAGAA-----AGGAGTACATGGGATCAAGATGTTGATC 57070			
Db 796 GlUlySHISGInaATrgInUleGInGInUlyATrgInUlySMetGInUlyThrhGIn 812			
QY 571 TTGTCTCAGAAATATTCGCCCACTGGAGCCCAAG-----GTGAAAGGTTACA 618			
Db 813 CysaSnAatrgATrhSerGInIleGInaIaGInPheGInSerAspCysGInUlySaIThr 8322			
QY 619 AAGGAAAGATTTGACGATATATCAACTGGAGGAATTCAAAGCCAGCTGGCTTCGG 678			
Db 833 GlUatrgCysGInSerAlaLeuGInSerLeuGInUlyATrgUlyAtrGIn----- 848			
QY 679 GAAATGATGTGCACAAAGGTTGTGGAAGATGCGTATACAGTGAATTAACCAACATG 738			
Db 849 -----GInUleuLySaPheGInGInGInaATrgInUlySerGInTrPhe 866			
QY 739 CAGAGAGATGAGCAAGAAAGAGCACAGAGATTCAGAGCAAAACTTAACAGAGATCTT 788			
Db 867 GlUlySaPrgInUleThrhGInGInUlySaIaGInaIaGInUleuLySGInUlyThrhLeu 886			
QY 799 GAA-----ATTAAGATCTGAGGAATATGAGAAATTTGAGAAATTTGAA 837			
Db 887 LysATrgGInUlySHIThRhSerLeuValLeuThrhGInGInUlyATrgInUlyMetLeuGIn 904			
QY 838 CTGAGATGAACAACAACACTTGTGACAGAGAGACAGAGCAAGAGCGCTTGGCCAGACAG 8977			
Db 905 ---LysThITrgUlySGInUlyLeu-----AsnSerMetValaIaGInaATrgGIn 919			
QY 898 GAGTGGCTGACA---CTAACGAAAGCTGGGGCAATCTGAGCAACCAACTGCACCTCAC 954			
Db 920 GInLeuLeuGInaPheGInUlySaPheUatrgaSnValaSerGInUlyThrhGInGInSerLeu 939			

```

1 OY      .955  AGAATCTAAATTAGCTCAACTCACTCAAGAAAAAGGTAATACATTAATGATTAATTGGGAAG  1014
2         :::::|||||
3 Db      940  SeraspelintleuleugtuleuylsSerSerHisLys----- 951
4 OY      1015  TTACAGAGAAGAAATGAAATGAGTGAAGAACACTGTGTCCACATGCGGAGACTACATAG  1074
5         ||| ||| ||||| ||| |||
6 Db      952  -----ArgtuleuAeuArgtuAtrgtuAluValleuCySglAlaEgYlaSerCluGln  969
7 OY      1075  ACGATGAAGCAAAAGGCTAAGGCACTGTGAT--AACCAACAGCCAGGCCACAGCCACAGC  1133
8         ||||| ||||| ||||| ||||| |||
9 Db      970  LeuAlaSerGlnArgtuGlnArgtuArgtuGlnuMetGlnHisAspGlnGluArgtuGlnuMet  989
10 OY      1132  CTGTGTGAGCTCTCTCAACAGCAGCAAC-----CAGCTTCTCGT  1170
11         :::::|||||
12 Db      990  MetSerLysLeuAlaMetGlnuAsnLleHisLysAlaThrCysGluThrAlaAspArg  1009
13 OY      1171  GAGAGGCAAGAGCCGTGTGCGAAGAGTGGAACGGCTCGGAGCCAGTACCAAGCAATGCCA  1230
14         ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
15 Db      1010  GlnArgtuAlaGluMetSerThrGluLleSerArgtuGlnSerLysIleLysGluMetGln  1029
16 OY      1231  CAATCT  1236
17         |||:::
18 Db      1030  GlnAla  1031
19
20 RESULT 2
21 US-09-727-384-6
22 : Sequence 6, Application US/09727384
23 : Patent No. US20020098511A1
24 :
25 : GENERAL INFORMATION:
26 :
27 : APPLICANT: Myriad Genetics, Inc.
28 :
29 : APPLICANT: Heichman, Karen
30 :
31 : APPLICANT: Cimbroa, Daniel M.
32 :
33 : APPLICANT: Bush, Angie
34 :
35 : APPLICANT: Mauck, Kimberly
36 :
37 : APPLICANT: Bartel, Paul L.
38 :
39 : TITLE OR INVENTION: Protein-Protein Interactions
40 :
41 : FILE REFERENCE: 2318-271
42 :
43 : CURRENT APPLICATION NUMBER: US/09/727,384
44 :
45 : CURRENT FILING DATE: 2000-12-01
46 :
47 : PRIOR APPLICATION NUMBER: US 60/168,377
48 :
49 : PRIOR FILING DATE: 1999-12-02
50 :
51 : PRIOR APPLICATION NUMBER: US 60/168,379
52 :
53 : PRIOR FILING DATE: 1999-12-02
54 :
55 : PRIOR APPLICATION NUMBER: US 60/185,056
56 :
57 : PRIOR FILING DATE: 2000-02-25
58 :
59 : NUMBER OF SEQ ID NOS: 8
60 :
61 : SOFTWARE: PatentIn version 3.0
62 :
63 : SEQ ID NO 6
64 :
65 : LENGTH: 2139
66 :
67 : TYPE: PRT
68 :
69 : ORGANISM: Homo sapiens
70 :
71 : US-09-727-384-6

```

	Alignment Scores:			
	Pred. No.:	3,21e-10	Length:	2139
	Score:	242.50	Matches:	107
	Percent Similarity:	41.86%	Conservative:	78
	Best Local Similarity:	24.21%	Mismatches:	158
	Query Match:	8.88%	Indels:	99
	DB:	10	Gaps:	18
US-09-502-945-1 (1-1552) x US-09-727-384-6 (1-2139)				
QY	10 GCATCCGAGAAAGCTAAACTTACTTATGAGAAAGCTGTAATTTAGAGATCCCAATTC	69		
	:::: :::: :::	::::		
Db	656 A1Ac1ng1u5nMc1y5G1n1aYh15G1u5nMc1u1h1aYh1r1eUg1u1y5G1n1l1e	675		
QY	70 AAGTTTGTGAGAGACGACTTACGTGATATACAGACACTTGTCAGATCTTAAAGACAA	129		
	: :			
Db	676 Ser1sp1euY5a5nG1u1l1e1aG1u1l1e1nG1n1c1y1n1a1l1a1v1a1l1eU5G1u---	694		
QY	130 CTAAGCATAAAGATTTCTTCTGGCTGCTAA1rACTTGTAAACCGTGTGGCTTTTGT	189		


```
Db 694 ----- 694
QY 190 TTGAATGCTGCTACAGTGAAGCTGTTTCCCAACCCATTAATGTTCAATATGACG 249
Db 695 -----AlaHisHisGluAlaThrCysArgHisGluGluGluValGluGluLeuGln 711
QY 250 ACCATCGAAGACCTGGTTAAAGAAAGATGACCTGATGCTGCACCTA----- 297
Db 712 Val-----LysLeuGluGluGluGluLysThrHisLeuGluGluLysLeuArgLeuGlnHis 729
QY 298 ---GTTTCGTAAGACGACGCTGGCAGATACGACGAAAGAGAACGATGCTTATGAA 354
Db 730 GluMetGluLeuValAlaArgLeuThrGlnAlaGln-----AlaSerPheGlu 745
QY 355 CAGGTGAACAAAGTTTCCAAATATCTGAGAGACCAATTTGAAAAAACCAAGCTTTA 414
Db 746 ArgGluArgGluGluGluLeuGlnSerSerAlaTrpThrGlu---GluLysValArgGluLeu 764
QY 415 ATCCAGTGTGACGAGTGTAGAGAGGCTGGAGAGCGGAGCGGACCTTGAAAAAGAA 474
Db 765 ThrGln-----GluLeuGlu-----GlnPheHisGluGluGln 775
QY 475 CTTCGATCTCAGACAGACAGAAAGGCCATTTGAGAAAGCATGATGAAAGAAATAACG 534
Db 776 LeuThrSerLeuValGluLysHisThrLeuGlnLysGluGluLeuArgLysGluLeuLeu 795
QY 535 AAAGAA-----AGGAGTACATGGGATCAAAAGATGTTGATC 570
Db 796 GluLysHisGlnArgGluLeuGlnGluGluLysArgGluLysMetGluThrGlu----- 812
QY 571 TTGCTCGAATATGTCACCACTGACGAGCCAG-----GTGAAAAAGGTTACA 618
Db 813 CysAsnArgArgThrSerGlnHisLeuAlaGlnPheGlnSerArgPcyGlnLysValThr 832
QY 619 AAGGAAAAAGATTCACTATTAATCACTGAGAGAAATCAAGCCAGCTGGCTTCGCG 678
Db 833 GluArgCysGluSerAlaLeuGlnSerLeuGluGluArgTyrArgGln----- 848
QY 679 GAAATGATGTCACAAAGGTGTGGAGAAATCGCCCTTCAGTCGTAATAAACCAACATC 738
Db 849 -----GluLeuLysAspLeuGlnGluGlnGluArgGluGluLysSerGlnTrpGluPhe 866
QY 739 GAGAGAGTGAAGCAGAAAGAGCAGACAGAGTTCAGAGCAAAACATAACAGGATCTT 798
Db 867 GluLysAspGluLeuThrThrSerLeuValLeuThrGlnGluArgGluMetLeuGln----- 904
QY 799 GAA-----ATTAAGATCAGAAATAGAGAAATTAGAAATAGAA 837
Db 887 LysArgGluLysThrThrSerLeuValLeuThrGlnGluArgGluMetLeuGln----- 904
QY 838 CTGATGTAAGACCAACTGTGGACACAGAGCGAGAGGCCCTGGCCAGAGAG 897
Db 905 ---LysThrTyrLysGlnHisLeu-----AsnSerMetValValGluArgGln 919
QY 898 GAGTGCCCTGAGA---CTAACAGAACTGCGGCGAATCTGAGACCACTCACCTCACCC 954
Db 920 GlnLeuLeuGlnAspLeuGlnAspLeuArgAsnValSerGluThrGlnGlnSerLeuLeu 939
QY 955 AGATCTGAATAGCTCAACTAGTCAGAGAAAAAGGTTACATATGATTAATTTGGGAAAG 1014
Db 940 SerAspGlnIleLeuGlnLeuLysSerSerHisLys----- 951
QY 1015 TTACAGAGAAAGAAATGAAGATTGGAGAGACAGTGTCCAGCATGGAGAGTACATAG 1074
Db 952 -----ArgGluLeuArgGluArgGluGluValLeuCysGlnAlaGluAlaSerGluGln 969
QY 1075 ACGATGAAGCAAGAGCTTAAGCGCTGGAT---AAGCAGACAGAGCCAGACCCAGAG 1131
Db 970 LeuAlaSerGlnArgLeuGlnArgLeuGlnMetGluHisAspGlnGluArgGlnGlnMet 989
QY 1132 CTGCTGACGCTCTCAGCAGACAGAAC-----CAGCTTCTCTCTG 1170
Db 990 MetSerLysLeuLeuAlaMetGluAsnIleHisLysAlaThrCysGluThrAlaAspArg 1009
```

```
QY 1171 GAGAGCAGAGCCTCTCGAAGAGGTGAGACCGGCTGGCCAGCCCACTTAACCAAGATGCCA 1230
Db 1010 GluArgAlaGluMetSerThrGluLeuSerArgLeuGlnSerLysIleLysGluMetGln 1029
QY 1231 CAATCT 1236
Db 1030 GlnAla 1031

RESULT 3
US-09-884-001-19
: Sequence 19, Application US/09884001
: Publication No. US20020182656A1
: GENERAL INFORMATION:
: APPLICANT: Bird, Timothy A.
: APPLICANT: Peschon, Jacques J.
: APPLICANT: Sims, John E.
: APPLICANT: Virca, G. Duke
: APPLICANT: Willis, Cynthia R.
: TITLE OF INVENTION: Methods for Regulating Vascularization Using GEF
: TITLE OF INVENTION: Containing NEK-Like Kinase (GNK)
: FILE REFERENCE: Immunex GNK/SGNK PCT
: CURRENT FILING DATE: 2001-06-18
: PRIOR APPLICATION NUMBER: US/09/884,001
: PRIOR FILING DATE: 1998-12-18
: NUMBER OF SEQ ID NOS: 19
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 19
: LENGTH: 868
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-884-001-19

Alignment Scores:
Pred. No.: 5,34e-10 Length: 868
Score: 238.50 Matches: 123
Percent Similarity: 39.08% Conservative: 99
Best Local Similarity: 21.65% Mismatches: 198
Query Match: 8.73% Indels: 148
Gaps: 18

US-09-502-945-1 (1-1552) x US-09-884-001-19 (1-868)
QY 37 GAGGAAAGTGTGAATGAGAGAAATCCCAATTGAAGTTTGAGAGAACTTAAGCTGAA 96
Db 27 GlnGluLysGluGluIleGlnLys---LysLeuSerGluSerArgHisGlnGlnAla 45
QY 97 TATCAGAACTGTGAGAGATCTTAAAGAGCAACTAAAGCAATTAAGAAATTTCTTGCGCT 156
Db 46 AlaThrThrGlnLeuGlnGlnLeuHisGlnGlnAlaLysArgGlnGlnGluValLeuAla 65
QY 157 GCTAATACTTAAACCGTGTGGTGCTGTTGTTGAAATGCTCAGCAATGAAGCTGTT 216
Db 66 -----ArgAlaValGlnGlnLysGluAla 73
QY 217 CTT-----TCCCAACCCACTAATGTTTCATATGACGACCAATCGAAGAA----- 261
Db 74 LeuValArgGluLysAlaAlaLeuGlnValArgLeuGlnAlaValGluArgAspArgGln 93
QY 262 ---CTGTTAAAGAAAGAGATGACTGTGATGCTGCACTAGTTTCCGTAAAGAGCAGCTTG 318
Db 94 AspLeuAlaGlnGlnLeuGlnGlnLysLeuSerSerAlaLysGluLeuLeuGlnSerLeu 113
QY 319 GCAGATACGACAGCAAGA-----GAA 339
Db 114 PheGluAlaGlnGlnGlnAlaAsnSerValIleGluValThrLysGlnLeuGlnValGln 133
QY 340 GCMAAGCTTTTGAACAGAGTAAACAACTTTTGCA-----ATATCTGAG 384
Db 134 IleGlnThrValThrGlnAlaLysGlnIleGlnGlnGluValArgCysLeuLysLeu 153
QY 385 GAAGCAATTTTGAAGAAACCAAGCT-----TTA 414
```

```

Db 154 GlnLeuAspThrGlnArgSerGlnAlaGlnGlnLysArgPalaAlaAlaArgGlnLeu 173
QY 415 ATCCAGTGTGACCAAGTTGAGAG---GAGCTGGAGAGCGAGCGGACCACTTGAAAA 471
Db 174 AlaGlnAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 193
QY 472 GAACCTGATCATCAGCAAGAGAAAGGCGCATTCAGCAAAACATGATGAAAGAAATA 531
Db 194 GluValAsnGlnLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 213
QY 532 AGCAAA-----GAAAGGAGTACATGGATCAAGATGATGATGCTGCT 576
Db 214 AlaLysAlaLeuGlnLysLeuGlnArgGlnLysMetGlnLeuGlnLysArgGln 233
QY 577 CAGATATTT----- 585
Db 234 GlnGlnThrGlnMetGlnAlaGlnAlaGlnArgGlnGlnGlnGlnGlnGlnGln 253
QY 586 -----GCCCACTGAGCGCCAGTGGAGAAAGTTACA----- 618
Db 254 SerAlaLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 273
QY 619 -----AAGCAAAAGATTTCAGCTATTATCACTGAGAGAAATTCAAAGC 663
Db 274 LeuLeuGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 293
QY 664 CAGCTGCTTCGCGGAAATGATGTCACAAAGTGTGTGAGAAATGCGCTATCAGCTG 723
Db 294 AspMetLysValGlnLysLeuLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 313
QY 724 AATTAACAACATGAGAGAGATGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 780
Db 314 GlnGlnAlaGlnArgGlnLysGlnAlaAlaAlaArgGlnGlnGlnGlnGlnGln 333
QY 781 -----AAACTAACAGGATCTTGAATTAAGATCAG 813
Db 334 LeuGlnGlnLysSerSerLeuLeuGlnAspLysMetSerLeuGlnLysGlnVal 353
QY 814 GAATATGAGAAATTTGAAATAGAACTGATGAAGCAACACTTTGAAAGAGAGAGAG 873
Db 354 AspLeuLysSerGlnLeuValAlaGlnAspSerGlnArgLeuValGlnGlnGln 373
QY 874 CAGAGGACAGCCCTGCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 933
Db 374 GlnGlnLysLeuArgGlnLysThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 393
QY 934 GAGCAGCACTGCGACCTGACCAATCTGAATAGCTCAA-----CTCAGTCMA 981
Db 394 LysAlaSerLeuThrLeuSerLeuMetGlnLysGlnGlnArgLeuValLeuGln 413
QY 982 GAAAAAAGTATCATATGATAATTTGGAAAGTTCAGAGAGAAATGAAAGATTGAG 1041
Db 414 AlaAspSerIleArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 433
QY 1042 GAACATGTGTCCAGATGGAGAGATACAGACATGACAGCAAGAGAGAGAGAGAG 1101
Db 434 GlnGlnGlnLysGlnLysLeuSerAlaGlnMetGlnLeuLeuArgGlnGlnLys 453
QY 1102 GAT-----AAGCAGCAGCCAGGCC 1119
Db 454 GluAlaAspPheLeuAlaGlnGlnAlaGlnLeuLeuGlnGlnLysLeuGlnAla 473
QY 1120 ACAGCCAGCAGCTGTGAGCTCTC-----AGCAAGAGAGAGAGAGAGAGAGAG 1167
Db 474 ThrGlnGlnGlnLeuArgAlaSerLeuThrPalaGlnGlnAlaAlaGlnLeuGln 493
QY 1168 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1206
Db 494 LeuArgLeuArgSerThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 513
QY 1207 -----CGAGCCAGTTACCCAGATGCCACATCTGATTCCTCAGCTGGA 1251

```

```

Db 514 AsnGlnAlaGlnAlaGlnAlaGlnLeuAlaSerLeu-TyrSerAlaLeu----- 539
QY 1252 TGGAACGAGTGAATTAATGAATTTACAAAGAGATATTTCATTCACTGTTAGACT 1311
Db 530 -----GlnGlnAlaLeuGlyse 535
QY 1312 AATATGCCACAGCAGCAGCAGCAGCTTCC---CAGGGTACAGCGGCTCAGCTGAG 1368
Db 535 ValGlnGlnSerArgProGlnLeuSerGlnGlnGlnGlnGlnGlnGlnGlnGln 555
QY 1369 GCGTGTCTCTATCAACGCGGG 1390
Db 555 YLeuGlnProAspGlnAsnGly 562

RESULT 4
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Fafis, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc. feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 3774181CD1
US-09-919-172-29

Alignment Scores:
Pred. No.: 2,13e-09 Length: 2125
Score: 231.50 Matches: 108
Percent Similarity: 41.20% Conservative: 98
Best Local Similarity: 21.60% Mismatches: 153
Query Match: 8.47% Indels: 141
DB: Gaps: 20

US-09-502-945-1 (1-1552) x US-09-919-172-29 (1-2125)
QY 37 GAGGAAAGGTGTAATTTGAGAAATCCCAATTGAAGTT-----TTGAGGAACGAC 87
Db 628 LysGlnLysSerArgValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 647
QY 88 TTAGCTGAATATAGAGAACTTGTGAAGATCTTAAGAGCA----- 129
Db 648 LeuArgLysGlnGlnArgGlnValGlnAspIleSerLeuGlnLysIleArgAlaGln 667
QY 130 -----CTAAGCATTAAGAAATTTCTTCG 153
Db 668 GluAlaLysGlnTyrArgArgGlnLeuGlnThrIleValArgGlnLysGln-----Ala 685
QY 154 GCTGCTAATATCTTGAACGCTGTGGTGTCTTTGTAATGTCTGACATGAAGCT 213
Db 686 AlaGlnArgGlnLeuGlnLysArgValArgGlnLeuThrIleGlnAlaGlnAlaLys 705
QY 214 GTTCTTCCCAACCCACTACTATGTTTCAATATCAG----- 249
Db 706 AlaValGlnLysLeuLeuAsnPheArgGlnGlnLeuGlnLysAsnThrPheThrArg 725
QY 250 ---ACCATGAAAGAGCTGTTAAAGAAAGAGATGATGATGCTGACACTTTCGCTA 306
Db 726 ArgThrLeuGlnLysPheLysLeuLysArgLysAsp----- 736
QY 307 AGGAGCAGCTTGGCAGATGCGCAGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 366

```

```

Db      1074 GlusnrlleValleuLgUlysgInPthrlleGInglInargCysGluaLeuLysIleIn 1093

RESULT 5
US-09-925-299-988
: Sequence 988, Application US/09925299
: Publication No. US20030040617A9
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
: FILE REFERENCE: PA102
: CURRENT APPLICATION NUMBER: US/09/925,299
: PRIORITY FILING DATE: 2001-08-10
: PRIOR APPLICATION NUMBER: PCY/US00/05883
: PRIOR FILING DATE: 2000-03-08
: PRIOR APPLICATION NUMBER: 60/124,270
: PRIOR FILING DATE: 1999-03-12
: NUMBER OF SEQ ID NOS: 1556
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 988
: LENGTH: 909
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (32)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (41)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (47)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (48)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (52)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (58)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (62)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (125)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (632)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (851)
: OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
: US-09-925-299-988

Alignment Scores:
Pred. No.: 2,53e-09 Length: 909
Score: 229.50 Matches: 100
Percent Similarity: 41.638 Conservative: 89
Best Local Similarity: 22.038 Mismatches: 196
Query Match: 8.40% Indels: 69
DB: 9 Gaps: 13

US-09-502-945-1 (1-1552) x US-09-925-299-988 (1-909)

QY      4 CTGGATGCATCGAGAGCTAAAACTTACTTATGAGGAAAGCTGTAATTGAGCAATCC 63
      |||::: ||||| ||| ::::| |::: |||
Db      354 LeuGluGluLysGluLysLeuLeuAlaThrGluGluGluAlaValAlaLysSer 373

QY      64 CAATTGAACTTTTGGAGGAACGACTTACTGTAATTCAGAGAACTTGT----- 111
      ::::| |::: |::: |::: |::: |::: |:::
Db      374 LysLeuArgGluLeuLeuAlaGluLysAlaLysAlaValAlaGlyGlu 393

```



```
QY 4 CTGATGATCCGAGAACTTAACTTATGAGAAAGTGAATTGAGATCC 63
    ||||| ||||| ||| :||| :|
Db 354 LeuGluGluysGluLysLeuLeuAlaThrGluGluInGluAspAlaValAlaLysSer 373
QY 64 CAATTGAATTTTGGAGAACGACTTACGATATACAGAACTTGT- 111
    :|||: ||| :|||:
Db 374 LysLeuArgGluLeuAsnLysGluMetAlaGluLysAlaLysAlaLacGlu 393
QY 112 GAAGATCTTAAGAGACTTAAGCATTAAGATTTCTTGGCTGCTTAATCTTGTAC 171
    :|||: ||||| :|||:
Db 394 AlaLysValLysGluLeuValAlaArgGluGluInGluLeuAlaValGluAlaArg 413
QY 172 CGTGTGTGTGCTTTGTTGTAATGTGTCACGATGAAGCTGTTCTTCCCAACCAT 231
    :|||: ||||| :|||:
Db 414 -----MetGlnAlaSerTyrArgGluHisValLysGluValGluIn 427
QY 232 ACTAATGTCATATGACACCATCGAA----- 258
    :|||: ||||| :|||:
Db 428 LeuGluGluLysIleArgThrLeuGluGluInGluLeuGluAsnGlyProAsnThrGluLeu 447
QY 259 ---AGACTGGTTAAAGAAAGATGACTTGTATGTCGACTAGTTTCCGTAAGAGACAGC 315
    ||||| :|||: |||
Db 448 AlaArgLeuGluGluInGluAsnSerIleLeuArgAspAlaLeuAsnGluAlaThrSerGlu 467
QY 316 TTGGCAGATACGACGAAAGCAAGCAAGTGTATGAAACGCTGAACAGTTTTCGA 375
    :|||: ||||| :|||:
Db 468 ValGluSerLysGluHisAlaGluLeuAlaLysLeuArgGluIn-----GluLeuSerLys 485
QY 376 ATATCTGAGAAAGCCAAATTTTGAAGAAACCAAGGCTTAAATCCAGTGTACAGTTGAGG 435
    :|||: ||||| :|||:
Db 486 ValSerLysGluLeu---ValGluLysSerGluAlaValAlaArgGluAsnGluGluInArg 504
QY 436 AAGAGCTGGAGAGCGGAGCGGAGCTTGAAGAAAGAA---CTTGCACTCTCAGCAAGAG 492
    ||| ||||| :|||: |||
Db 505 LysAlaLeuGluAlaLysAlaAlaPheGluLysGluValLeuGluInGluAlaSer 524
QY 493 AAAAGGGCCATGAGAAAGACATGATGTAAGAA-----GAAATPACGAAAGAAAGGAG 546
    ||| ||||| :|||: |||
Db 525 HisArgGluSerGluGluAlaLeuGluInLysArgLeuAspGluValSerArgGluIn----- 542
QY 547 TACATGGATCAAGATGTTGATCTTGTCTCAGAAATATGTCGCCCACTGAGAGCCGAGGTG 606
    :|||: ||||| :|||:
Db 543 -----LeuGluHisThrGlnSerHisAlaSerLeuArgAlaAspAla 557
QY 607 GAAAAGCTTACAAAGAAAGATTTACGCTTAACTCACTGAGAGAAATTCAAAGCCAG 666
    ||||| :|||: |||||
Db 558 GluLysAlaGluGluInGluIn-----GluGlnMetAlaGluLeuHisSerLys 573
QY 667 CTGCTCTCTCGGAAATGATGTCACAAAGGTGTGTGAGAA-----ATGGGCTAT 717
    ||| ||| ||||| :|||:
Db 574 LeuGlnSerSerGluAlaGluValArgSerLysCysGluGluLeuSerGluLeuHisGly 593
QY 718 CAGCTGATATAAACCAACATGAGAGAGATGAGCGCAAGAAAGAGACAGAGATTCAGA 777
    ||||| :|||: |||||
Db 594 GlnLeuGlnGluAlaArgAlaGluAsnSerGlnLeuThrGluArgGluLeuArgSerIleGlu 613
QY 778 GCAAAATCTACAGGAGATCTGAAATTTAAAGATCAGAAATAGAGAAATTTGAGATAGAA 837
    ||| :|||: ||||| :|||:
Db 614 AlaLeuLeuGluAlaGlyGlnAlaArgAspAlaGlnAspValGlnAlaSerGln**Glu 633
QY 838 CTGATGAAACCAACCAACACTTGGACACGAGCAGCAAGAGGACCCCTGGCCAGAGAG 897
    ||||| :|||: |||||
Db 634 AlaAspGlnGluInThrArgLeuLysGluLeuGlnSerGlnValSerGluLeuGluLys 653
QY 898 GAGTGCCCTGAGACTAAGCAAGCTGTGGCGAATCTGAGACCAACTGACCTCAGCCAGAG 957
    ||| :|||: ||||| :|||:
Db 654 GluAlaIleGluLeuArgGluAlaValGluGluInLysValLysAsnAsnAspLeuArg 673
QY 958 TCTGAA-----ATAGCTCACTCAGTCAAGAAAAAGTATATACATATGATATAA 1005
    :|||: ||||| :|||:
Db 674 GluLysAsnTrpLysAlaMetGluAlaLeuAlaThrAlaGluGluInAlaCysLysGluLys 693
```

```
QY 1006 TTGGGAAAGTTACAGAGAAAGAAATGAAATTTGAGAGACAG---TGTGCCAGCATGG 1062
    ||| :|||: ||||| :|||:
Db 694 LeuHisSerLeuThrGlnAlaLysGluLysGluSerGluLysGlnLeuCysLeuIleGluAla 713
QY 1063 AGAGTACATGAGACGATAGACCAAGGCTTAAGCAGCTGAT----- 1104
    :|||: ||||| :|||:
Db 714 GlnThrMetGluAlaLeuLeuAlaLeuLeuProGluLeuSerValLeuAlaGlnGlnAsn 733
QY 1105 -----AACACAGACCCAG 1116
    :|||: ||||| :|||:
Db 734 TyrThrGluTrpLeuGlnAspLeuLysGluLysGlyProThrLeuLeuLysHisProPro 753
QY 1117 GCCACAGCCCGACAGCTGTGACGCTCTCCACAGACAGACAGACTTCTCTGAGAGAG 1176
    ||| :|||: ||||| :|||:
Db 754 AlaProAlaGluProSerSerAspLeuAlaSerLysLeuAlaArgGluAlaGluGluThrGln 773
QY 1177 CAGAGCCCTGTGGAGAGGTGACCGGCTGCGGACCCAGTTA 1218
    :|||: ||||| :|||:
Db 774 SerThrLeuGlnAlaGluCysAspIleTyrArgSerIleLeu 787

RESULT 7
US-09-820-843A-113
: Sequence 113, Application US/09820843A
: Publication No. US20030039963A1
: GENERAL INFORMATION:
: APPLICANT: Council of Scientific and Industrial Research
: TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
: FILE REFERENCE: Q63915
: CURRENT FILING DATE: 2001-03-30
: NUMBER OF SEQ ID NOS: 118
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 113
: LENGTH: 2354
: TYPE: PRM
: ORGANISM: L. major
: NAME/KEY: misc_feature
: OTHER INFORMATION: AC005802_5 L6202.3
: NAME/KEY: misc_feature
: OTHER INFORMATION: g16899670
US-09-820-843A-113

Alignment Scores:
Pred. No.: 3.33e-09 Length: 2354
Score: 229.00 Matches: 95
Percent Similarity: 43.18% Conservative: 98
Best Local Similarity: 21.25% Mismatches: 198
Query Match: 8.38% Indels: 56
DB: Gaps: 14

US-09-502-945-1 (1-1552) x US-09-820-843A-113 (1-2354)
QY 7 GATGCATCCGAGAACTTAACTTATGAGAAAGTGAATTGAGAAATCCCA 66
    :|||: ||||| :|||: |||||
Db 872 GluLeuGlnGlnArgLeuAspThrAlaThrGlnGlnAlaGluLeu---GluAlaGln 890
QY 67 TTGAAGTTTGGAGAACGACTTACGATATACAGAACTTGAAGATCTTAAGAG 126
    :|||: ||||| :|||:
Db 891 ValAlaArgLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAspThrAlaThrGln 910
QY 127 CACTTAAGCATTAAGATTTCTTGTGCTGCTAATCTTGAACCGTGTGTGCTCT 186
    ||| ||| ||||| :|||:
Db 911 GlnArgAlaGluLeuGluIn-----AlaArgValAlaArgLeu 922
QY 187 TGTGTAATGTGCTCAGCATGAAGCTGTCTTCCCAACCAACTACTATGATTCAT--- 243
    ||| :|||: ||||| :|||:
Db 923 -----AlaAlaAspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGlu 939
QY 244 ---ATGACAGCCATCGAAAGACTGTGAAGAAAGATGACTTGATCTGACACTGTT 300
    ||| :|||: ||||| :|||:
Db 940 LeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGluAlaGlnLeuAla 959
```

```
QY 301 TCCGTAAGGACGAGCTTGGAGATACGACAGCAAAAGACCAAGTCTTATGACAGT 360
   ::::: ::::: ::::: ::::: :::::
Db 960 ArgLeuAlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeu 979
QY 361 AACAAGCTTTGCAATATCTGAGCAAGCCAAATTTGAAAAAACCAGGCTTTATCCAG 420
   ::::: ||| ::::: ::::: ||| :::::
Db 980 GlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuAlaGlnLeuAlaArg 999
QY 421 TGT-----GACCAGTTGAGGAAGAGAGCTGAGAGCGAGCGAGCACTTGA 468
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 1000 LeuAlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGln 1019
QY 469 AAA-----GAACTTGATCTCAGCAAGAAAAGG 498
   ::::: ||::: ||::: ||::: ||:::
Db 1020 GlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuAlaGlnGlnValAlaArgLeu 1039
QY 499 GCCATTGAGAAACATGATGAAAAAGAAATACCAAGAAAGGAGTCAATGGGATCA 558
   ||| ::::: ||| ::::: ||| :::::
Db 1040 AlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGln 1059
QY 559 AAGATGTTGATCTTCTCAGAAATATGCCCAGTGGAGCGGAGTGAAGATTACA 618
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 1060 ArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAlaArgLeuAla 1079
QY 619 AAGGAAAAGATTTTCACTATTATCAACTG-----GAGAAATTTCAAAGCCAG 666
   ::::: ||| ::::: ||::: ||::: ||:::
Db 1080 AlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArg 1099
QY 667 CAG-----GCTTCTCGGGAATGATGTCACAAAGGCTGTGCA 705
   ||| ||| ||| ||| |||
Db 1100 LeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnValAlaArgLeuAlaAla 1119
QY 706 -----GAAATGCGCTATCAGCTG-----AATAAACCAACTGGAAG----- 744
   ||| ||| ||| ||| |||
Db 1120 AspGlyAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArgLeu 1139
QY 745 GATGAGCGCAAAAGAGACAGAGCTTCAGAGCAAAAACCTACAGGATCTTGAAT 804
   ||| ||| ::::: ||| ::::: |||
Db 1140 AspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAlaArgLeuAlaAlaAsp 1159
QY 805 AAGAGCAAGAAATAGCAAAATAGCAATAGCAATGAGATGGAAGCAAAACCTGGAA 864
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 1160 ArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArgLeuAsp 1179
QY 865 CAGGACAGCAGAGAGCGCTGCGCAGAGAGAGTGCCTGAGACTTAACAGAACTGCTG 924
   ||::: ||| ::::: ||| ::::: |||
Db 1180 ThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAspArg 1199
QY 925 GCGGAATCTGAGACCACTGCACTGCCACCAAGATCTGAATAGCTCAA----- 972
   ||::: ||::: ||::: ||::: ||:::
Db 1200 AspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArgLeuAspThr 1219
QY 973 CTCAGTCAGAAAAAAGATATCATATGTAATTTGGGAAAGTTACAGAGCAAAATGAA 1032
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 1220 AlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnValAlaArgValAlaAspGlyAsp 1239
QY 1033 GAATTTGAGAAACAGTGTCTCCAGATGGAGAGTACATGAGAGCTGAAGCAAGAGCTA 1092
   ||| ::::: ||| ::::: ||| ::::: |||
Db 1240 GlnAlaArgGlnGlnLeuAlaAlaAsnAla-----GlnGluLeuGlnGlnArgLeu 1256
QY 1093 AGGCACTGATAGACACAGCCAGGCGCAGCCAGCCAGCAGCTGCTCAGCAAG 1152
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 1257 AspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAsp 1276
QY 1153 CAGAACACACTTCTCTCGAGAGGCGAGAGCTGTCCGAAGAGGTGACGCGCTGCGGACC 1212
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 1277 ArgAspGlu-----AlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGln 1293
QY 1213 CAGTTACCAGCATGCCCAA 1233
   ::::: |||
Db 1294 ArgLeuAspThrAlaThrGln 1300
```

```
RESULT 8
US-09-991-496-120
; Sequence 120, Application US/09991496
; Patent No. US20020169285A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probst, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
; FILE REFERENCE: 210121.420C9
; CURRENT APPLICATION NUMBER: US/09/991,496
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120
; LENGTH: 2310
; TYPE: PRT
; ORGANISM: Leishmania major and chagasi
US-09-991-496-120

Alignment Scores:
Pred. No.: 3,95e-09 Length: 2310
Score: 97 Matches: 97
Percent Similarity: 43.418 Conservative: 104
Best Local Similarity: 20.958 Mismatches: 202
Query Match: 8.35% Indels: 60
DB: Gaps: 14

US-09-502-945-1 (1-1552) x US-09-991-496-120 (1-2310)
QY 7 GATGATCCGAGAGAGCTTAACCTTACTTATGAGAAAAGTGTGAATTTAG----- 57
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 493 GlnLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgVal 512
QY 58 -----GAATCCCAATGTAAGTTTGGAGAACGAC 87
   ||| ||| ||| ||| |||
Db 513 AlaArgLeuAlaAlaAspArgSpGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGlu 532
QY 88 TTAGCTGAATATCAGAGACTTGTGAAGATCTTAAGAGCAACTAAAGCATTAAGAATTT 147
   ||| ::::: ||| ::::: ||| ::::: |||
Db 533 LeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAla 552
QY 148 CTTCTGCTGCTAATACT-----TGTAACCGTGTGGGCTGCTTTGTTGAATGT 198
   ||| ||| ||| ||| |||
Db 553 ArgLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArg 572
QY 199 GCTCAGCATGAAAGCTGTTCTTCCCAAGCCACTACTAATCTTCAT-----ATGCAGAC 252
   ||::: ||| ||::: ||| ||::: |||
Db 573 AlaGlnLeuGlnAlaGlnValAlaArgLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArg 592
QY 253 ATCGAAGACTGGTTAAGAAAGAGATGACTGATGCTGCATCAATTTCCGTAAGAGAGC 312
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 593 LeuAspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaArgValAlaArgLeuAlaAla 612
QY 313 AGCTTGGCAGATACGAGCAAGAGCAAGAGTCTTATGACAGGTGAACAAAGTTTG 372
   ::::: ||::: ||::: ||::: ||::: ||:::
Db 613 AspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArgLeu 632
QY 373 CAATATCTGAGAGACCAATTTGAAAAAACCAAGCTTTAATCCAGTGT----- 423
   ::::: ||| ::::: ||| ::::: |||
Db 633 AspThrAlaThrGlnGlnArgAlaGlnLeuGlnAlaGlnLeuAlaArgLeuAlaAlaAsp 652
QY 424 ---GACCAGTTGAGGAAGAGCTGAGAGGAGCGGAGCGAGCTTGAAGAAAGACTT--- 477
   ||::: ||::: ||::: ||::: ||::: ||:::
Db 653 GlyAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGlnGluLeuGlnGlnArgLeuAsp 672
```

```
QY 478 ---GCATCTCAGCAAGAG-----AAAAGGCCATTGAGAA 510
      |||:|||||
Db 673 ThrAlaThrGlnGlnInArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArg 692
QY 511 GACATGATGAAAGAAATTAACGAAAGAAAGGAGTACATGGGATCAAGATGTTGATC 570
      |||:|||||
Db 693 AspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThr 712
QY 571 TTGCTCAGATATATGCCCCCACTGAGCCCGAGTGGAAAAAGTTACAAAGCAAAAGATT 630
      |||:|||||
Db 713 AlaThrGlnGlnInArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArg 732
QY 631 TCACCTATTATCACTG-----GAGGAATTCACAGCCAGCTG----- 669
      |||:|||||
Db 733 GluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThrAla 752
QY 670 -----GCTTCTCGGAAATGAGTGTCAAAAGGTGTGGA-----GAA 708
      |||:|||||
Db 753 ThrGlnGlnInArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspGlu 772
QY 709 ATGGCGTATGAGCTG-----AATAAACCAACATGGAGAG-----GATGAGCGAGAA 756
      |||:|||||
Db 773 AlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThrAlaThr 792
QY 757 AAGGAGCAGAGAGATTCCAGAGCAAAACTTAACAGGATCTTGAATTAAGATCAGAGAA 816
      |||:|||||
Db 793 GlnGlnInArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArgAspGluAla 812
QY 817 ATAGAAATTTGAGATTAAGAACTGGATGAAGCAACACACTTGGAAACAGAGCAGCAG 876
      |||:|||||
Db 813 ArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThrAlaThrGln 832
QY 877 AAGGCAAGCCCTGGCGCAAGAGAGAGTGCCTGAGACTAACAGACTGCGGCGGATCTGAG 936
      |||:|||||
Db 833 GlnInArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAspGluAlaArg 852
QY 937 CACCAACTGCACTCACCAGATCTGAATAGCTCAA-----CTCAGTCAAGAA 984
      |||:|||||
Db 853 GlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThrAlaThrGlnGln 872
QY 985 AAAAGGTATCATTTGATTAATTTGGAAAGTTACAGAGAAAGAAATTAAGAAATTTGGAGAA 1044
      |||:|||||
Db 873 ArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArgAspGluAlaArgGln 892
QY 1045 CAGGTGTCCAGATGGGAGAGTACGAGCAGTGAAGCAAGCAAGCAAGGACACTGGAT 1104
      |||:|||||
Db 893 GlnLeuAlaAlaAsnAla-----GlnGluLeuGlnInArgLeuAspThrAlaThr 909
QY 1105 AAGCAGACAGAGCCAGAGCCAGCTGCTGAGCTCTCAGCAAGCAGAAACAGCTT 1164
      |||:|||||
Db 910 GlnGlnInArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAspArgAspGlu 928
QY 1165 CTCTGAGAGAGCAGACCTGTCTGGAAGAGGTGAGCCGGCTCGGACCCACTTACCCAGC 1224
      |||:|||||
Db 929 -----AlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThr 946
QY 1225 ATGCCACAA 1233
      |||
Db 947 AlaThrGln 949
```

```
RESULT 9
; Sequence 114, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
; FILE REFERENCE: 063915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
```

```
; SEQ ID NO 114
; LENGTH: 2310
; TYPE: PRF
; ORGANISM: L. major
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: AC005893_12 L6202.3
; NAME/KEY: misc_feature
; OTHER INFORMATION: 916899664
; US-09-820-843A-114

Alignment Scores:
Pred. No.: 3,95e-09 Length: 2310
Score: 228.00 Matches: 97
Percent Similarity: 43.41% Conservative: 104
Best Local Similarity: 20.95% Mismatches: 202
Query Match: 8.35% Indels: 60
DB: Gaps: 14

US-09-502-945-1 (1-1552) x US-09-820-843A-114 (1-2310)
QY 7 GATCATCCGAGAAAGCTTAATCTTATGAGGAAAGTGTGAATTGAG----- 57
      |||:|||||
Db 493 GluLeuGlnInArgLeuAspThrAlaThrGlnGlnInArgAlaGluLeuGluAlaArgVal 512
QY 58 -----GAATCCCAATTGAGGTTTGGAGGAGAC 87
      |||:|||||
Db 513 AlaArgLeuAlaAlaAspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGlu 532
QY 88 TTACCTGATATTCAGAGAACTTGTGAAGATCTTAAAGAGCAACTTAAGCATTAAGAAATTT 147
      |||:|||||
Db 533 LeuGlnInArgLeuAspThrAlaThrGlnGlnInArgAlaGluLeuGluAlaArgValAla 552
QY 148 CTTCTGCTGCTTAATCT-----TTTACCCTGCTGCTGCTTTGTTGAATGT 198
      |||:|||||
Db 553 ArgLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThrAlaThrGlnGlnArg 572
QY 199 GCTGAGATGAGCTGTTCTTCCCAACCATCTTAATGTTCAAT-----ATGCAAGCC 252
      |||:|||||
Db 573 AlaGluLeuGluAlaGlnValAlaArgLeuAlaAlaAsnAlaGluGluLeuGlnInArg 592
QY 253 ATCGAAAGACTGTTTAAAGAAAGATGACTGTGTCGTGCACTTAAGTTCCTGAAGGAGC 312
      |||:|||||
Db 593 LeuAspThrAlaThrGlnGlnInArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAla 612
QY 313 AGCTTGCAGATPACGACGCAAGAGCAAGCTTATGACAGGTTGAAGGAGTTTG 372
      |||:|||||
Db 613 AspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeu 632
QY 373 CAATATCTGAGGAAGCAATTTGAAACCAAGGCTTTAAATCCAGTGT----- 423
      |||:|||||
Db 633 AspThrAlaThrGlnGlnInArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAsp 652
QY 424 ---GACCATTTGAGGAAGAGCTGAGAGCGGAGCGCACTTGAAGAAAGATT-- 477
      |||:|||||
Db 653 GlnArgGlnInArgGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeu 672
QY 478 ---GCATCTCAGCAAGAG-----AAAAGGCCATTGAGAA 510
      |||:|||||
Db 673 ThrAlaThrGlnGlnInArgAlaGluLeuGluAlaArgValAlaArgLeuAlaAlaAspArg 692
QY 511 GACATGATGAAAGAAATTAACGAAAGAAAGGAGTACATGGGATCAAGATGTTGATC 570
      |||:|||||
Db 693 AspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThr 712
QY 571 TTGCTCAGATATATGCCCACTGAGCCCGAGTGGAAAAAGTTACAAAGCAAAAGATT 630
      |||:|||||
Db 713 AlaThrGlnGlnInArgAlaGluLeuGluAlaGlnLeuAlaArgLeuAlaAlaAspArg 732
QY 631 TCACCTATTATCACTG-----GAGGAATTCACAGCCAGCTG----- 669
      |||:|||||
Db 733 GluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnInArgLeuAspThrAla 752
```

```

QY 670 -----GCTTCTGGGAATGATGTCACAAAGCTGTGCA-----GAA 708
Db 753 ThrGlnGlnArgAlaGluLeuGlnLysAlaArgLeuAlaAlaAspGlyAspGlu 772
QY 709 ATGGCGTATCAGCTG-----AATAAACCAACATGAGAG-----GATGAGCGAGAA 756
Db 773 AlaArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAspThrAlaThr 792
QY 757 AAGGACACAGAGACTTCAGAGCAAAACTAACAGGATCTTGAATTAAGATCAGAA 816
Db 793 GlnGlnArgAlaGluLeuGlnLysAlaArgValAlaArgLeuAlaAlaAspArgAspGluAla 812
QY 817 ATAGCAAAATTGAAATGAACTGATGTAAGCAAACTTGAACAGCAGCAG 876
Db 813 ArgGlnGlnLeuAlaAlaAsnAlaGluGluLeuGlnGlnArgLeuAspThrAlaThrGln 832
QY 877 AAGGACCGCTGGCCGACAGAGAGTCCGAGACTAACAGAACTGCTGGCCGAACTGAG 936
Db 833 GlnArgAlaGluLeuGlnLysAlaArgLeuAlaAlaAspGlyAspGluAlaArg 852
QY 937 CACCACTGCACTCACCAGATCTGAAATAGCTCAA-----CTCAGTCAAGAA 984
Db 853 GlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThrGlnGln 872
QY 985 AAAAGGTATACATGATGATTAATGGAAAGTTACAGAGAAAGAAATGAGAGAA 1044
Db 873 ArgAlaGlnLeuGlnLysAlaArgValAlaArgLeuAlaAlaAspArgAspGluAlaArg 892
QY 1045 CAGTGTGTCCACATGAGAGAGTACATGACAGATGACAAAGCTAAGGACGCTGAGT 1104
Db 893 GlnLeuAlaAlaAsnAla-----GlnGluLeuGlnGlnArgLeuAspThrAlaThr 909
QY 1105 AAGCAGACGACGACGACGACGACGCTGTGACAGCTCTCAGCAGAGCAACGACTT 1164
Db 910 GlnGlnArgAlaGluLeuGlnLysAlaArgLeuAlaAlaAspArgAspGlu---- 928
QY 1165 CTCCTGGAGAGCAGACCTGTGCGAGAGGTGACCGGCTGGCGCAAGTTACCCAGC 1224
Db 929 -----AlaArgGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThr 946
QY 1225 ATGCACCAA 1233
Db 947 AlaThrGln 949

RESULT 10
US-09-874-923-120
: Sequence 120, Application US/09874923
: Patent No. US20020081320A1
: GENERAL INFORMATION:
: APPLICANT: Reed, Steven G.
: APPLICANT: Campos-Neto, Antonio
: APPLICANT: Webb, John R.
: APPLICANT: Dillon, David C.
: APPLICANT: Skelky, Yasir A.W.
: APPLICANT: Bhatia, Ajay
: APPLICANT: Colier, Rhea
: APPLICANT: Probst, Peter
: APPLICANT: Brannon, Mark
: TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
: TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF LEISHMANIASIS
: FILE REFERENCE: 210121.420C8
: CURRENT APPLICATION NUMBER: US/09/874.923
: NUMBER OF SEQ ID NOS: 122
: SOFTWARE: SeqID for Windows Version 4.0
: SEQ ID NO 120
: LENGTH: 2310
: TYPE: PRT
: ORGANISM: Leishmania major and chagasi
US-09-874-923-120
Alignment Scores: 3.95e-09 Length: 2310

```

```

Score: 228.00 Matches: 97
Percent Similarity: 43.41% Conservative: 104
Best Local Similarity: 20.95% Mismatches: 202
Query Match: 8.35% Indels: 60
Db: 10 Gaps: 14

US-09-502-945-1 (1-1552) x US-09-874-923-120 (1-2310)
QY 7 GATCATCCGACGACGATAAATCTTACTTATGAGAAAGTGTAATGAG----- 57
Db 493 GlnLeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGlnLysAlaArgVal 512
QY 58 -----GAATCCCAATTGAAAGTTTGGAGCAAGCAG 87
Db 513 AlaArgLeuAlaAlaAspArgAspGluAlaArgGlnGlnLeuAlaAlaAsnAlaGluGln 532
QY 88 TTAGCTGATATCAGAGAACTTGTGAAGATCTTAAGACCAATAAGCATTAAGAAATT 147
Db 533 LeuGlnGlnArgLeuAspThrAlaThrGlnGlnArgAlaGluLeuGlnLysAlaArgValAla 552
QY 148 CTTCTGCTGTCTATCT-----TGTAACCGTGTGCTGCTTTGTTGAATGT 198
Db 553 ArgLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThrGlnGlnArg 572
QY 199 GCTCAGCATGAAGCTGCTTCTCCCAACCCATCTAATGTTCTAT-----ATGCAGACC 252
Db 573 AlaGlnGlnLysAlaGlnLysAlaArgLeuAlaAlaAsnAlaGluGlnGlnArg 592
QY 253 ATCGAAGACTGTTAAAGAAAGAGTACTGATGCTGACACTATTCGTAAGAGAC 312
Db 593 LeuAspThrAlaThrGlnGlnGlnArgAlaGluLeuGlnLysAlaArgValAlaArgLeuAlaAla 612
QY 313 ACCTTGACATACGACGACGACGACGACGCTTATGACAGCTGTAAGCAAGTTTGT 372
Db 613 AspArgAspGluAlaArgGlnGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeu 632
QY 373 CAATATCTGAGAGACCAATTTGAAAAAACCAAGCTTTAATCCAGTGT----- 423
Db 633 AspThrAlaThrGlnGlnGlnArgAlaGluGlnLysAlaArgLeuAlaAlaAsp 652
QY 424 ----GACCAGTTGAGAGAGAGAGCTGGAGAGCGGCGGACGACTGAAAAAGAACTT 477
Db 653 GlyAspGluAlaArgGlnGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAsp 672
QY 478 ----GCATCTCAGCAGAG-----AAAAAGGCCATTGAGAA 510
Db 673 ThrAlaThrGlnGlnArgAlaGluLeuGlnLysAlaArgValAlaArgLeuAlaAlaAspArg 692
QY 511 GACATGATGAAGAAAGAAATTAACGAAAGAAAGGAGTACATGGGATCAAGATGTTGATC 570
Db 693 AspGluAlaArgGlnGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThr 712
QY 571 TTGTCTCAGAAATATGCCCACACTGGGCGGCGGAGGAAAGGTTTCAAGGAAAGAT 630
Db 713 AlaThrGlnGlnArgAlaGluLeuGlnLysAlaArgLeuAlaAlaAspArgAsp 732
QY 631 TCAGCTATTATCAACTG-----GAGAAATTCAAAGCCAGCTG----- 669
Db 733 GlnAlaArgGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAla 752
QY 670 -----GCTTCTGGGAATGATGTCACAAAGCTGTGCA-----GAA 708
Db 753 ThrGlnGlnArgAlaGluLeuGlnLysAlaArgValAlaArgLeuAlaAlaAspGlyAspGlu 772
QY 709 ATGGCGTATCAGCTG-----AATAAACCAACATGAGAG-----GATGAGCGAGAA 756
Db 773 AlaArgGlnGlnLeuAlaAlaAsnAlaGluGlnGlnArgLeuAspThrAlaThr 792
QY 757 AAGGACACAGAGACTTCAGAGCAAAACTAACAGGATCTTGAATTAAGATCAGAA 816
Db 793 GlnGlnArgAlaGluLeuGlnLysAlaArgValAlaArgLeuAlaAlaAspArgAspGluAla 812
QY 817 ATAGCAAAATTGAAATGAACTGATGTAAGCAAACTTGAACAGCAGCAG 876

```


[illegible]

```

Db 731 ArgArgArgGluAspAsnAspSerSerGlnLeuGlnValGlnGluLeuLeu 750
QY 964 ATAGCT-----CAACTGAGTCAAGAAAAAGGTATATGATATGATTAATGGAAAGTTA 1017
Db 751 MetAlaMetGluLysValLysGlnGlnLeuGlnSerMetLysAlaLysLeuSerSerThr 770
QY 1018 CAGAGAGAAATGAGAAATTTGGAGAACAGTGTCTCAGCATGGAGAGATGATGAGACG 1077
Db 771 GlnGlnSerLeuAlaGlnLysGln-----ThrHisLeuThr 782
QY 1078 -----ATGACGAAGAGCTTAAGGAGCTGGATAGCACAGCCAGCCACAGCCAGCAG 1131
Db 783 AsnLeuArgAlaGlnLysArgLysHisLeuGlnGlnValLeuGlnMetLysGlnGlnAla 802
QY 1132 CTGGTGCACCTCTCTCAGCAAGCAG-----AACCACTTCTCTCTGAGAGAGCAGAC--- 1182
Db 803 LeuLeuAlaAlaIleSerGlnLysAspAlaAsnIleAlaLeuLeuGlnLeuSerSerSer 822
QY 1183 -----CTGTCCGAGAGAGGTGGAGCGGCTGCGGACG 1212
Db 823 LysLysLysThrGlnGlnGlnValAlaIleAlaLeuLysArgLysLysAspArgLeuValGln 842
QY 1213 CAGTTACCCAGCATGCCACAATCTGATTGCTGACCTGATGAGACAGAGTGAATTAATG 1272
Db 843 GlnLeuLysGlnGlnThrGln-----AsnArgMetLysLeuMet 855
QY 1273 -----ATTACAAACAG 1284
Db 856 AlaAspAsnTyrGlnAsp 861

RESULT 12
US-09-820-843A-73
; Sequence 73, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 73
; LENGTH: 1805
; TYPE: PRT
; ORGANISM: M. genitalium
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: g1|1045905
US-09-820-843A-73

Alignment Scores:
Pred. No.: 6,86e-09 Length: 1805
Score: 224.50 Matches: 98
Percent Similarity: 43.10% Conservative: 83
Best Local Similarity: 23.33% Mismatches: 166
Query Match: 8.22% Indels: 73
DB: 9 Gaps: 15

US-09-502-945-1 (1-1552) x US-09-820-843A-73 (1-1805)
QY 16 GAGAGCTAAACTTACTTATGAGAAAAGTGAATTTGAGGAATCCCAATTGAAGTTT 75
Db 495 GlnGlnLeuTyrLeuValLysLysGlnLysGlnAspGlnLysGlnLeuLeuLeuPhe 514
QY 76 TTGAGAGACACTTACCTGAATATTCAGAGAACTTGAAGATCTTAAGACAGCAATGAAG 135
Db 515 PheGlnLysGlnLeuLysGlnHisGlnAlaAspPheGlnAsn---GlnLeuGlnAlaLys 533
QY 136 CATAAAGATTTCTTCTGCTGCTCTAATACTTGTAAACCGCTTGCTGCTTTGTTGAAA 195

```

```

Db 534 GlnGlnGlnLeuPheGlnAlaLysHisAlaLeuGluArg-----SerPheIleLys 550
QY 196 TGTGTCACATGATGACTGTCTTTTCCCAAAACCATTATATGATATGACAGACATC 255
Db 551 LeuGlnAspLysGlnLysAspLeu-----AsnThrLysAlaGlnGlnIle 565
QY 256 GAAAGACTGGTTAAAGAAAGATGACTTGATGCTGACATCTTCCGTAAGACAGAC 315
Db 566 Ala-----AsnGlnPheSerGlnLeuLysThrLysPheLysSerLys 578
QY 316 TTGGCAGATACGACGCAAGAGACAGACAGTCTTATGACAGAGTGAACAATTGGCAA 375
Db 579 SerAlaAspPheGlnLeuMetLeuGlnAsnGlnLysGlnLysLeuGlnGlnGlnLysGln 598
QY 376 -----ATATTCGAGGAAGCAATTTTGAATAAACCAAGCCTTAATC-----CAGTGT 423
Db 599 LysLeuPheGlnGlnLysArgThrTyrPheGlnLysArgAsnAlaIleValLeuSerAsnAlaGln 618
QY 424 GACCACTTGAAGAGAGAGCTGGAGAGCAGCGAGCGAGCAGCTTGAAAGAACTTGACAT 483
Db 619 GlnGlnLysArgGlnGlnLeuLeuGlnGlnGlnLysGlnLysThrLeuAspGlnLeuThrLysSer 638
QY 484 CAGCAAGAAAAAGGCCATTTGAGAAAGACATGATGAAAAAG-----GAAATTAAG 534
Db 639 PheGlnGlnGlnLysArgLeuLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 658
QY 535 AAAGAAAGGAGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
Db 659 LysGlnLysGlnLysLeuGlnLysLysLeuGlnAspPheSerGlnThrSerLeuAsnAla 678
QY 574 TCTCAGAAATATGCCCCA-----CTGAGAGCCAGGTGGAAG 612
Db 679 SerLysAsnLeuAlaGlnLysArgLysGlnMetAlaIleLysPheLysGlnLysGlnGlnGlnGln 698
QY 613 GTTACAAAGAAAAATTCACGCTATTAATCACTGAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 672
Db 699 ThrGlnLysGlnLeuLeuAsnAspValAsnAsnAlaGlnValIleGlnAlaAspLeuAla 718
QY 673 TCTCGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 732
Db 719 -----GlnLeuAsnGlnSerLeuAsnGlnGln 727
QY 733 AACATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 792
Db 728 ArgSerGlnLeuGlnAsnAlaLysGlnArgLysAlaIleAspPheHisAsnAspSerLeuLys 747
QY 793 GATCTTGAATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 843
Db 748 LysLeuAsnGlnLysGlnLeuSerLeuGlnLysArgLeuGlnGlnLeuGlnGlnGlnGlnGlnGln 767
QY 844 ---GAAAGCAACACACACTTGGAAACAGAGCAGCAGAGCAGCAGCAGCAGCAGCAGCAGCAG 900
Db 768 AlaAsnGlnLysHisSerTyrGlnAsnGln-----AlaTyrPheGlnGlnGlnGln 784
QY 901 TGCCTGAGACTAACAGACTGCTGGCGCAATCTGACACACCACTGACACTGACCAATATCT 960
Db 785 -----LeuAspLysLeuAsnArgGlnLysGlnAlaPheLeuAsnLeuArgLysLys 801
QY 961 GAAATGATCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1005
Db 802 GlnThrMetGlnValAlaAspAlaIleLysGlnArgLeuSerAspLysHisGlnAlaLeuAsn 821
QY 1006 -----TTGGGAAAGTTTACAGAGAAAGTGAAGAAATTTGAGAGAACAGTGTCTCAGCAT 1059
Db 822 MetGlnGlnAlaGlnLeuAsnArgLysThrHisGlnLysLeuAsnAlaPheLeuAsnHis 841
QY 1060 GGGAGAGTACATGAGACGATGAAAGCAAGAGCTTAAGCAGCAGCTGATTAAGACAGCAGGCC 1119
Db 842 AspAlaAspGlnLysSerLeuGlnAspGlnLeu-----Ala 853
QY 1120 ACAGCCGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1179
Db 854 ThrValLysGlnThrGlnLysLeuIleAspLeuGlnArgSerAlaLeuLeuGlnLysGln 873

```


Db 594 ArgIleasn-----ProThrThrValLysMetLys 603
OY 1267 TAATGAAATTACAAGAGATATTACATTCATCTGGTTAGACTTAATATG 1317
 ::: ||| |||||::: ||| |||
Db 604 SerSerValPheaspGlnaspLysThrPheValAlaGlnThrLeuGlnMet 620

Search completed: March 21, 2003, 13:55:53
Job time : 47.6182 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 21, 2003, 12:34:32 ; Search time 12.0233 Seconds
(without alignments)
7595.986 Million cell updates/sec

Title: US-09-502-945-1

Perfect score: 2732
Sequence: 1 cttctgtagctccgagaa.....aaatgaactttaagaaga 1552.

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=slp
-Q=/cgn2_1/USPTO.spool/US09502945/runatc_14032003_101059_19145/app.query.fasta_1.10979
-DB=Issued_Patents_AA -QFMT=fastan -SUFFIX=ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsun62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUPMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09502945.eccn_1.1.80_@runatc_14032003_101059_19145 -KCPU=6 -ICPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	233.5	8.5	631	4	US-08-477-831C-11 Sequence 11, Appl1
2	232.5	8.5	976	4	US-09-104-324B-4 Sequence 6, Appl1
3	232.5	8.5	2482	1	US-08-328-254-6 Sequence 2, Appl1
4	232.5	8.5	606	4	US-08-477-831C-2 Sequence 1, Appl1
5	230.5	8.4	1939	4	US-09-310-187A-1 Sequence 1, Appl1
6	227.5	8.3	3248	5	US-08-353-700-1 Sequence 1, Appl1
7	227.5	8.3	3248	5	PCT-US95-16216-1 Sequence 1, Appl1
8	224	8.2	1388	4	US-09-572-191-2 Sequence 2, Appl1
9	224	8.2	1388	4	US-09-723-262-2 Sequence 2, Appl1
10	224	8.2	1388	4	US-09-723-219-2 Sequence 2, Appl1
11	223	8.2	1898	1	US-08-056-200-94 Sequence 94, Appl1
12	223	8.2	1898	2	US-08-800-644-94 Sequence 94, Appl1

13	221.5	8.1	1388	2	US-08-685-576-1 Sequence 1, Appl1
14	221	8.1	955	1	US-08-006-676B-1 Sequence 1, Appl1
15	221	8.1	955	1	US-08-282-845-2 Sequence 2, Appl1
16	221	8.1	955	2	US-08-428-414A-3 Sequence 3, Appl1
17	221	8.1	955	5	PCT-US94-00324-1 Sequence 1, Appl1
18	220.5	8.1	1886	4	US-08-938-105-3 Sequence 3, Appl1
19	218.5	8.0	2101	1	US-08-466-390-4 Sequence 4, Appl1
20	218.5	8.0	2101	1	US-08-470-950-4 Sequence 4, Appl1
21	218.5	8.0	2101	1	US-08-467-781-4 Sequence 4, Appl1
22	218.5	8.0	2101	1	US-08-195-487-4 Sequence 4, Appl1
23	218.5	8.0	2101	2	US-08-483-924-4 Sequence 4, Appl1
24	218.5	8.0	2101	4	US-09-452-294-1 Sequence 1, Appl1
25	218.5	8.0	2101	5	PCT-US93-06160-4 Sequence 4, Appl1
26	218	8.0	1093	5	PCT-US93-03077-1 Sequence 1, Appl1
27	217.5	8.0	1388	2	US-08-685-576-4 Sequence 4, Appl1
28	217	7.9	1312	2	US-08-687-080-51 Sequence 51, Appl1
29	216.5	7.9	885	2	US-08-533-306A-4 Sequence 4, Appl1
30	216.5	7.9	885	2	US-08-742-923A-4 Sequence 4, Appl1
31	216	7.9	1312	2	US-08-592-126-148 Sequence 148, App
32	215.5	7.9	1354	3	US-08-685-871-2 Sequence 2, Appl1
33	215	7.9	514	2	US-08-960-022-14 Sequence 14, Appl1
34	214.5	7.9	576	2	US-08-533-306A-2 Sequence 2, Appl1
35	214.5	7.9	576	2	US-08-742-923A-2 Sequence 2, Appl1
36	214.5	7.9	816	2	US-08-533-306A-6 Sequence 6, Appl1
37	214.5	7.9	816	2	US-08-742-923A-6 Sequence 6, Appl1
38	213	7.8	900	2	US-08-630-822A-62 Sequence 62, Appl1
39	213	7.8	900	2	US-09-005-069-62 Sequence 62, Appl1
40	213	7.8	900	4	US-09-171-156A-21 Sequence 21, Appl1
41	209.5	7.7	1066	4	US-09-541-782-8 Sequence 8, Appl1
42	209.5	7.7	1066	4	US-09-723-820-8 Sequence 8, Appl1
43	208	7.6	443	2	US-08-795-475-6 Sequence 6, Appl1
44	201.5	7.4	1162	2	US-08-728-323A-2 Sequence 2, Appl1
45	201.5	7.4	1162	4	US-09-298-568-2 Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-477-831C-11
; Sequence 11, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUMEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477, 831C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PIERRE, MARGARET A.
; REGISTRATION NUMBER: 30,709
; REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
;

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: /desc = "RHAMM I-2a"
US-08-477-831C-11

```

```

Alignment Scores:
Pred. No.: 1,286-13 Length: 631
Score: 233.50 Matches: 126
Percent Similarity: 41.47% Conservative: 100
Best Local Similarity: 23.12% Mismatches: 170
Query Match: 8.55% Indels: 149
DB: Gaps: 28

```

US-09-502-945-1 (1-1552) x US-08-477-831C-11 (1-631)

```

QY 16 GAGAGCTAAACTTACTTAAAGTGAATGAGAA----- 60
DB 58 GIULYSGLULYSLE-----ASPGIULYSCYSGLUTHRGILULYSLEULEUGLUTYRILE 75
QY 61 -----TCCCAATGAGATTTTGGAGACGACTTACGTGAATAT 99
DB 76 GINGLUILESERCYSALASERASPGIUNVALIGIULYSCYSLYSVALASPILEALAGINLEU 95
QY 100 CAGAGAACTTGGAAGTCTTAAAGACCAATTAAGCATTAAGATTTCCT 150
DB 96 GIU-----GIUASPLEULYSGLU-----LYSASPAIGIUILEUSERLEULYS 110
QY 151 -----CTGGCTGCTAAACT-----TCTAACCGTGTGGTGCTTGTGTTGAATGT 198
DB 111 GINSEULEUGLUGLUNASINLEUPHESERLYSGINLEGLUASPLEULHRVALLYSCYS 130
QY 159 -----GCTCAGCATGAGCTGTCTTCCCAAAACCAT----- 231
DB 131 GINLEULEUGLUTHRGILUARGSPASINLEUVALISERLYSASPAIGIULARGIAGLUTHR 150
QY 232 ACTAATGTCATATGCAGACATC---GAAAGACTGGTTAAAGAAAGAGATGAC----- 282
DB 151 LEUSERALAGLUMETGILNLEULHRGILUARGLEUALALEUGLILUARGLUGLUTYRGILU 170
QY 283 ---TGATGTCGCACTAGTTCCGTAAGCAGACTTGGCAGATACGACAGCAAGAGAA 339
DB 171 LYSLEUGLUGLULYSGLULENGINSEULEUL-----GLNGLINLUULYSGLU 188
QY 340 GCAGTGTCTTATGACAGGTGAACAAGTTTGCAGATATCTGAGAGAACCAATTTTGA 399
DB 189 LEUSERALIAARGLEUGLN---GLNGLINLEUCYSERPHGGLNGLINLEUHRSERGLU 207
QY 400 AAACCC-----AAGCTTTAATCCAGCTGTGACCAGTTGAGAGAG 438
DB 208 LYSASNVALPHELYSGLUGLULYSLEUALALEUALAGLULEUASPLAVALAGLINGLN 227
QY 439 GAGCTGAGAGGAGCGAGCGAGGACTTGAAGAAAGAACTTGCACTTCAGCAAGAGAAAG 498
DB 228 LYS---GLUGLUGLINSERGLUARGLEUVALIYLSGLINLEUGLUGLUGLILUARGLYSERTHR 246
QY 499 GCC-----ATTGAGAAAGACATGATGAATAAGAAAGAAAGAAAG 543
DB 247 ALAAGLUGLINSERLTHRARGLEUASPSINLEUARGGILULYSGLIUALGILULEUCLULYS 266
QY 544 GAGTACATGGCATCAAGATGTTGATCTTG-----TCT 576
DB 267 HISILEALALANIALAGLINALILEULEULEALAGLINGLULYSTYRASNAPTHRALA 286
QY 577 CAGATATATGCCCACTGGAGCCCGAGGTGGAAAGGTTTCAAGAGAAAGATTTCACGT 636
DB 287 GINSEULEUARGSPVALIHRALAGLINSERVALAGLINGLULYSTYRASNAPTHR 306
QY 637 ATTATCAACTGAGAGAAATTCAGACCGAGCTGCTCGGGA----- 681
DB 307 ALAGINSEULEUARGSPVALIHRALAGLINSERGLUGLUGLULYSTYRASNAP 326

```

```

QY 682 -----ATGATGTCAACAAGTGTGTGCAANAATGCGCTATACGTGAAT 726
DB 327 THRALAGINSEULEUARGSPVALIHRALAGLINSERGLUGLULYSTYRASNAP 346
QY 727 AAACCAACATGTGAG---AAGAT-----GAGCAGAAAGAGGACAGACA 768
DB 347 ASPTHRALAGLINSERLEUARGSPVALIHRALAGLINSERLVALGILULYSTYR 366
QY 769 GAGTTCAGAGCAAAAACTAACAGGATCTT----- 798
DB 367 ASNAPTHRALAGLINSERLEUARGSPVALIHRALAGLINSERLUSERLYRYSER 386
QY 799 -----GAAATTAAGATCAAGCAAAATGAGCAATGACATAGAA----- 837
DB 387 THRLEULYSGLULILEGLUASPLEULYSLEUGLUNASINLEUHRLEULGILULYSVALALA 406
QY 838 ---CTGATGAAAGCAAAACATCACTTGAACAGGAG----- 870
DB 407 METALAGIULYSERVALGILUNSPVALIGLINGLILILEUHRALAGLUSERTHRASN 426
QY 871 -----CAGCAAGAGCAGCCCTGGCCAGACAGAG 900
DB 427 GINGLUTYRALARGMETVALGILNASPLEUGLINASNARGSERTHRLEULYSGLULINGLU 446
QY 901 TGCCTGACACTAACA-----GACTGCTGGCGCAATCTGAGCACCACACTGCAC 948
DB 447 ILELYSGIUILETHRSESRPHEULEUGLULYSILETHRASPLEULYSANGLINLEUARG 466
QY 949 CTCACACAGATCGAATA---GCTCAGCTCAGTCAAGAAAGGATATCATATGATATA 1005
DB 467 GINGLNASPGLUASPHEARGLYSLEUGLUGLULYSGLULYSARGTHRILAGLULYS 486
QY 1006 TTGGGAAAGTTACAGAGAAAGAAATGAGAAATGAGAGAACTGTGCTCCAGCATGGAGAGA 1065
DB 487 GIUASNVALMETHRGILUENURHMETGILUILEANLYSTRPARGLEULENURYRASPGLU 506
QY 1066 GTACATGACACATGATAGCAAAAGCTAAGCGACGTGGAT----- 1104
DB 507 LEUTYRGILUTYTHRLYSPROPHEGLNGLINGLINSLEUASPLAPHEGLUALAGLULYSGLU 526
QY 1105 -----AAGCACAGCCAGCGCACGCCGACACTG----- 1134
DB 527 ALALEULASNGLINHISLEY---ALATHRGINGLUGLINSLEUASNLYSILLEARGASPSER 545
QY 1135 ---GTGACACTCTCAGACAGACAGAC-----CAGCTT 1164
DB 546 TYRALAGLINSLEUGLIYHISGLINASINLEULYSGLINLYSILEYSHISVALIYALSLEU 565
QY 1165 CTCCTGGAGAGGACAGACCTGTGCGAAGAGGTGACCGGCTCGGACCCAGTTACCACAC 1224
DB 566 LYSASPGILUASINSERGILNLEULYSERGLUVALISERLYSLEUARGSERGLINLEUVALYS 585
QY 1225 ATGCCCAATCTGAT 1239
DB 586 ARGLYSGILNASGLU 590

```

RESULT 2
US-09-104-324B-4
Sequence 4, Application US/09104324B
Patent No. 6232460
GENERAL INFORMATION:
APPLICANT: T recti, Ozlem; Sahin, Ugur; Pfeundschuh, Michael
TITLE OF INVENTION: Methods For Diagnosis And Treating Cancers,
TITLE OF INVENTION: And Methods For Identifying Pathogenic Markers In A Sample
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fulbright & Jaworski LLP
STREET: 666 Fifth Avenue
CITY: New York City
STATE: New York
ZIP: 10103
COMPUTER READABLE FORM:


```

MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage
COMPUTER: IBM
OPERATING SYSTEM: PC-DOS
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/104,324B
FILING DATE: 25-June-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/892,702
FILING DATE: 15-July-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hanson, No. 623246oman D.
REGISTRATION NUMBER: 30,946
REFERENCE/DOCKET NUMBER: LUD 5491
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 318-3000
TELEFAX: (212) 752-5958
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 976 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-09-104-324B-4

Alignment Scores:
Pred. No.: 1,91e-13 Length: 976
Score: 232.50 Matches: 100
Percent Similarity: 42.60% Conservative: 93
Best Local Similarity: 22.08% Mismatches: 147
Query Match: 8,51% Indels: 113
Gaps: 18

US-09-502-945-1 (1-1552) x US-09-104-324B-4 (1-976)
QY 16 GAGAGCTRAAACTTACTTATGAGAGAAAGTGTGAATTGAGGAATCCCAATTGAATT 75
   |||:||||| ||| :||| ||| ||| ||| |||
Db 388 GluGluLeuLeuArgThrGluGluGlnArgLeuGluLysAsnGluAspGlnLeuLysIle 407
QY 76 TTGAGGAAGCACTAGCTGAGAA----- 96
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 408 LeuThrMetGluLeuGlnLysLysSerSerGluLeuGluMetThrLysLeuThrAsn 427
QY 97 ---TATCAGAGAACTGTGCAGATCTTAAAGCAACTAAAGCAATTAATTTCTTCTG 153
   |||:||||| |||:||||| ||| ||| ||| ||| |||
Db 428 AsnLysGluValGlnLeuGlnGlnLeuLysValLeuGlnLysLeuThrLeuLeu 447
QY 154 GCTGCTAATCTTGTAAACCGTGTGCTGCTTTGTTGAATGTGCTCAGCATGAAGCT 213
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 448 TyrGluAsn-----LysGlnPheGluLys 455
QY 214 GTT-----CTTCCCAAGCCATACTAATGTTCATATGCAAGCAACCAATCGAAGAGCTG 264
   |||:||||| ||| ||| ||| ||| ||| ||| |||
Db 456 IleAlaGluGlnLeuLysGlnThrGlnGlnLeu---IleGlnLeuGlnIleAlaArg 474
QY 265 GTTAAAGAAAGAGATGACTTGTATGTCGCATAGTTCCGTAAAGAGCAGCTTGCGAGAT 324
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 475 GluLysGluValIleAspLeuGlnIleGlnLeuThrAlaIleThrSer----- 491
QY 325 ACGCAGCAAGAAAGAAAGTGCCTTATGACAGGTGAACAAAGTTTCCAAATATCTGAG 384
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 -----GluGlnTyrTyrSerLysGluValLysAspLeuLysThrGlnLeuGln 507
QY 385 GAAGCAATTTTGAAGAAACCAAGCTTAAATCCAGTGTGACCAATG-----AGG 435
   |||:||||| |||:||||| ||| ||| ||| ||| |||
Db 508 AsnGluLysLeuLysAsnThrGlnLeuThrSerHisCysAsnLysLeuSerLeuGlnAsn 527
QY 436 AAGAGAGTGAAGAGCGAGCGAGCTTGAAGAAAGAACTTGACTCAGCAAGAGAAA 495
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 528 LysGlnLeuThrGlnGlnThrSerAspMetThrLeuGlnLeuLysAsnGlnGlnIle--- 546
QY 496 AGGCCATTTGAGAAAGACATGATGAAAAAATACGAAAGAAAGGAGTACATGGGA 555
   ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 547 -----AspIleAsnAsnAsnLysLysGlnGlnGluArg----- 557
QY 556 TCAAAAGATGTGATCTTGTCTCAGAAATATGCCCCAAGTGGAGCCAGCTGCAAAAAGTT 615
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 -----MetLeuLysGlnIleGlnLysLeuGlnGlnThrGlnLeuArg----- 573
QY 616 ACAAGAGAAAGATTTACACTATTAATCAATGGAGGAAATCAAGCCAGCTGCTTCT 675
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 574 -----AsnGlnLeuGlnTyrValArgGlnLeuLysGln 585
QY 676 CGGAAATGATGTC-----ACAAGGTGTGTGAGCAAAATGCGC 714
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 LysArgAspGlnValLysCysLysLeuAspLysSerGlnGlnAsnCysAsnLeuArg 605
QY 715 TATCAGCTGCAATTAACCAACACTGGAGAGATGTAGGCGCAAGAAAGCAGCAGAGACTTC 774
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 LysGlnValGlnAsnLysAsnLysTyrIleGlnGlnLeuGlnGlnLysAsnLysAlaLeu 625
QY 775 AGAGCAAAA-----ACTAACAGGATCTTGAATTAAGATCAGGAATTAAGAAA 825
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 LysLysLysGlyThrAlaGlnSerLysGlnLeuAsnValTyrGlnIleLysValAsnLys 645
QY 826 TTGAGAAATAGAACTGAGTGAAGAAAGCAACAA----- 855
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 LeuGlnLeuGlnLeuGlnSerAlaLysGlnLysPheGlnLysIleThrAspThrTyrGln 665
QY 856 -----CACTTGCAACAGGAGCGAGCAAGGCA 882
Db 666 LysGlnIleGlnAspLysLysIleSerGlnGlnAsnLeuGlnGlnValGlnLysAla 685
QY 883 GCCCTGGCCAGAGAGAGTGCCTGAGACTA--ACAGAACTGCTGGCGCAATCTGAGCAC 939
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 LysValIleAlaAspGlnAlaValLysLeuGlnLysGlnLysAspLysArgCysGlnHis 705
QY 940 CAACTGCACCTCAGCAGATCTGAATTAAGCTCACTCACTGCAAGAAAAGTATACATAT 999
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 706 LysIle-----AlaGlnMetValAlaLeuMetGlnLysHisGlnTyr 721
QY 1000 GATTAATTTGGGAAGTTACAGAGAAAGAAATGAAGATTTGGAGAACAGTGTCTCCAGCAT 1059
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 722 AspLysIle-----IleGlnGlnLysArgSerGlnLeu----- 732
QY 1060 GCGAGATGATCATGAGCAGTGAAGCAAGGCTTAAGCAGCTGATTAAGCAGCAGCGCC 1119
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 ---GlyLeuTyrLysSerLysGlnGlnGlnSerSerLeuArgAlaSerLeuGlnIle 751
QY 1120 ACAGCCAGCAGCTG---GTGCACTCTCTCAGCAGCAGCAGCAGCAGCTTCTCTGAGAGG 1176
   ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 752 GluLeuSerAsnLeuLysAlaGlnLeuLeuSerValLysLysGlnLeuGlnIleGlnArg 771
QY 1177 CAGAGCCTGTGAGAGAGGTGAGCCGGCTGCGGAGCCAG 1215
Db 772 -----GluGlnLysGlnLysLeuLysArgGln 780

RESULT 3
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESS: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-328-254-6

Alignment Scores:
Pred. No.: 2,84e-13 Length: 2482
Score: 232.50 Matches: 119
Percent Similarity: 40.51% Conservative: 88
Best Local Similarity: 23.29% Mismatches: 185
Query Match: 8.51% Indels: 119
DB: 1 Gaps: 19

US-09-502-945-1 (1-1552) x US-08-328-254-6 (1-2482)
QY 1 CTTCTGATGATCCGAGAAAGCTAAACTTATTGAGAAAGTGAATTGGAA 60
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1542 IIELEASPLAIGLNASNLSERLYSALAGIUALGIUTHTLEULYSTRGIIIEGLUGIU 1561
QY 61 -----TCCCAATGAAGTTTTTGAGAAAGACTTGAATATTCAGAAACTTGTGA 114
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1562 METALAARGSERLEULYALPHGIIULEUASPLLEUVALTHTLEULYSTRGIIULYSGLIU 1581
QY 115 GATCTTAAAGACCACTAAAGCTAAAGAAATTTCTTGCGTCTGCT-----AATACTTGT 168
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1582 AANLEUHTHLYSGIIIEGLIULYSGLIULYGLNLEUSERGIULEUASPLYSLEULEU 1601
QY 169 AACCGGTGGGGCTTTGTTGAATGTGCTCAGCATGAA---GCTGTTTCCCAA 225
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1602 SERSEPHLYSERLEULEUGIULYGLIULYGLIULYGLIULYGLIULYGLIULYGLIU 1621
QY 226 ACCCATCTAATGTTTCATATGACAGACCATCGAAAGACTGGTTAAAGAAAGATGACTTG 285
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1622 SERLYSTHRLAIVAGIULWET-----LEUINASNGLILEULYSGIULEUASNGLIUAL 1639
QY 286 AIGTCGCACTAGTTCCTGAAGGACAGCTTGCGAGATACGCAAGCAAGCAAGCAAGT 345
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1640 VALAIALALEUCYSGIYASPSGLIULIEMELYSALATHTRGIIULINSEULEUASPR 1659
QY 346 GCTTATGACAGGTGAACAACAAGTTTGCAAATATCTGAGGAACCATTTTGAAGAAACC 405
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1660 PROIIEGLIUGIULHISGLINLEU-----ARGASNLSERIEGLIULYSLEU 1674
QY 406 AAGGCTTTAATCCAGTGTGACCAAGTTGAGG-----AAGGAG 441
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1675 ARGALIAARGLEUGIULNASPSGLIULYSGLINLEUCYVALLEUGINLEULYSGIU 1694
QY 442 CTGGAGAGCGACGCG-----GAGCGACTTGAAGAAAGCACTT--- 477
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1695 SERGIUHNHISALASPLEULEULYARGVALGIIULASNLEUGIULARGIULEUGIU 1714
QY 478 -----GCATCTCAGCAAGAAAGGCCATT----- 504
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1715 IIEALIAARGTHASNGIULHISALALALEUGIUALAGIULASNLSERLYSGIULIUAL 1734

```

```

QY 504 ----- 504
DB 1735 GIUTHTLEULYSALALYSTIEGLIUGIMETHICLSERLEUARGGLYLEUGIULEUASPR 1754
QY 505 -----GAGAAAGACATGATGAAAAAGAAATTAACGAAAGAGGAG 546
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1755 VALIATHTRLIARGSERGIULYSGLIULASNLEUHTHASNGLILEULYSGIULIUGIU 1774
QY 547 TACATGGGATCAAAAGATGTTGATCTGTCT-----CAGAAATTTGCCCCAAGCGAGGCC 600
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1775 ARGIIIESERGIULEUGIULIETLEIASNSESRPHEGLIULASNILEUGIULYSGIU 1794
QY 601 CAGGTGAAAAAGGTTCAAAAAGAAATTTACAGTAT-----AATCAAA 645
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1795 GIIUGIULYSVALGIMETLYSGIULYSSESRTHALAMETGLIMETLYSGIULIUGIU 1814
QY 646 CTGGAGAAATTCAGAACAGCTGCTGTCTGGGAAATGAGTCAAAAGTGTGT--- 702
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1815 LEULYSGIULIULASNGLIULARGVALAIALALEUHNIAASNAPR---GIIUGIUALCYSLYS 1833
QY 703 -----GAGAAATGGCGCTTCAAGTGAATTAACCAACATGGAGAG----- 744
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1834 ALALYSGIULIULASNLEUSERSERGIULVALGICYLEUGIULEUGIULYSALAGIULIU 1853
QY 745 -----GATGAGCGAGAAAGAGACACAGAGAGTTCAGACCAAAACTAAC--- 789
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1854 LEUGIULYULNASPSGLIULALYSASNANTYRLIETALLEUGINSESRVALASNGLIU 1873
QY 790 -----AGGATCTTGAATTAAGATCAAGAAATPAGAG 822
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1874 LEUIIEGLIULIUALIULASPSGLIULYSGLIULYSLASPSGLIULIIESESR 1893
QY 823 AANTGAGAAATAGACTGTGATGAGAAACCAACAA-----CACTTGAACAGAGACAG 873
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1894 ARGLEULYASNSGLIIEGLIASNPSGLIULIULYVALSERLYSLEUSERSGLIULIUGIU 1913
QY 874 CAGAGAGCAGCCCTGGCCAGAGAGAGTGCCTGAGACTTAACAGAACTGCTGGCGAAATCT 933
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1914 GLIULIHNISGLIULIULYSTRPYSGLIULIULYSLASPSGLIULIULYVALGIULIU 1933
QY 934 GAGCACCAGCACTGCACCTCAACAGATGTGAATAGCTCAACTGCTCAAGAAAAAGGTAT 993
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1934 GIUGIULYSLIEGLIULIULIULYGLNLSERLYSASNALISERLEUIN----- 1948
QY 994 ACATATGATTAATTTGGGAAAGTTACAGAGAAGAAATGGAATTTGGAGCAACAGTGT--- 1050
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1949 -----ASPTHTLEULIULIULYGLINSESRTYRLIYASNSLEULIULASNGLIULIU 1966
QY 1051 -----GTCCAGCATGGGAGAGTACATGAG 1074
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1967 LEUTHTLYMETASPRYSMETSESRPHEVALGIULYVALASNLYSMETHRLATALYSGLIU 1986
QY 1075 ACGATGAACCAAGGCTGATGAGCGAGTGAATACACAGCAGCGCCACCGAGCAGCTG 1134
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1987 THHGLILEUGIN-----ARGIULMETHNISGLIULIULYSTRHIAAGIULIU 1204
QY 1135 GTGCACTCTCTCAGCAAGCAAGCAAGCAAGTTCCTCTGAGAGGAGCGCTGTGCAAGAG 1194
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2005 GIUGIULYSESRGIULYULYASNARGLEUALIAGIULIULIULYSEULIULIUGIU 1204
QY 1195 GTGGACCGCGCTGCGGACCAAGTTACCCAGCATG 1227
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2025 IIEYSESRSELYSASPSGLIULIULYSGIULIU 2035
; : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-08-477-831C-2
; Sequence 2, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUWEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN

```

TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
 NUMBER OF SEQUENCES: 60
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: FISH & NEAVE
 STREET: 1251 AVENUE OF THE AMERICAS
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10020-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Rel. #1.0, ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US-08/477,831C
 FILING DATE: 07-JUN-1995
 CLASSIFICATION: 514
 ATTORNEY/AGENT INFORMATION:
 NAME: PIERRI, MARGARET A.
 REGISTRATION NUMBER: 30,709
 REFERENCE/DOCKET NUMBER: STM-10
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-596-9000
 TELEFAX: 212-596-9090
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 606 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 DESCRIPTION: /desc = "RHAMM I protein"
 US-08-477-831C-2

Alignment Scores:
 Pred. No.: 1,74e-13 Length: 606
 Score: 232.00 Matches: 120
 Percent Similarity: 41.51% Conservative: 100
 Best local Similarity: 22.64% Mismatches: 174
 Query Match: 8.49% Indels: 136
 DB: 4 Gaps: 25

US-09-502-945-1 (1-1552) x US-08-477-831C-2 (1-606)

QY 4 CTGATGATCCGAGAGCTAAACTTACTATGAGAAAGTGTGAATTTAGGAATCC 63
 DB 54 LeucysAlaSerAspGlnVal- - - - -GluysCysLysValAspIleAla 68
 QY 64 CAATGAAGTTTGGAGAGCACTTACGTAATATCAGAGACTGTGAGATCTTAAA 123
 DB 69 GlnLeuGlu- - - - -GluAspLeuLysGlnLysAspArgGluIleLeuSerLeuLys 85
 QY 124 GAGCACTAAAGCATAAAGAAATTTCTTGTGCTGTAATCTTAAACGTTGGTGT 183
 DB 86 GlnSerLeuGlnGluAsn- - - - -IleThrPheSerLysGlnIleGluAsp 100
 QY 184 CTTTGTTGAATGT- - - - -GCTCAGCATGAGCTGTTCTTCCCAAAACCAT 231
 DB 101 LeuThrValLysCysGlnLeuLeuGlnThrGluArgAspAsnLeuValSerLysAspArg 120
 QY 232 - - - - -ACTAATGTCATATGACAGACATC- - -GAAAGACGTGTTAAAGAA 273
 DB 121 GluArgAlaGlnThrLeuSerAlaGlnMetGlnIleLeuThrGluArgLeuAlaLeuGln 140
 QY 274 AGAGATGAC- - - - -TTGATGCTCTGACACTAGTTTCCGTAAGAGACGCTTGACAGAT 324
 DB 141 ArgGlnGluArgGlnLysLeuGlnGlnLysLeuGlnSerGlnSerLeuLeu- - - - - 158
 QY 325 ACGGAGCAAAAGAGCAAGCTTATGACAGAGGTGAACAGAGTTTGCAATATCTGAG 384
 DB 159 GlnGlnGlnLysGlnLeuSerAlaArgLeuGln- - -GlnGlnLeuCysSerPheGlnGln 177

QY 385 GAACCAATTTGAAAAACC- - - - -AAGGCTTAATCCAGTGT 423
 DB 178 GlnMetThrSerGlnLysAsnValPheLysGlnGluLeuLysLeuAlaLeuAlaGlnLeu 197
 QY 424 GACCAGTTGAGGAAGAGCTGGAGAGCGGAGCGACTTGAAAAAGAACTTGACATCT 483
 DB 198 AspAlaValGlnGlnLys- - -GlnGlnGlnSerGlnArgLeuValLysGlnLeuGlnGln 216
 QY 484 CAGCAGAGGAAAGGGCC- - - - -ATTGGAAGAAAGCAATCATGATAAAGGAA 528
 DB 217 GluArgLysSerThrAlaGlnGlnLeuThrArgLeuAspAsnLeuLeuArgGlnLysGln 236
 QY 529 ATAAGGAAGGAAGGAGTACATGAGTCAAGATGTTGATCTTG- - - - - 573
 DB 237 ValGlnLeuGlnLysHisIleAlaIleHisIleAlaIleLeuIleAlaGlnGlnLys 256
 QY 574 - - - - -TCTCAGAAATATTGCCACTGGAGCCCGAGGTGGAAGGTTACAAG 621
 DB 257 TyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSerValGlnGln 276
 QY 622 GAAAGATTTCAGCTATTATCACTGAGGAAATTCAAAGCCAGCTGCTTCGGGAA 681
 DB 277 LysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSerGlnGln 296
 QY 682 - - - - -ATGATGTCACAAGAGTGTGTGAGAAATG 711
 DB 297 GluLysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSerGln 316
 QY 712 CGCTATCAGCTGAATTAACCAACATGAG- - -AAGAT- - - - -GAGGCA 753
 DB 317 GlnGlnLysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGlnSer 336
 QY 754 GAAAGGAGCAGACAGAGATTGAGCAAAACTAAGAGGATCTT- - - - - 798
 DB 337 ValGlnGlnLysTyrAsnAspThrAlaGlnSerLeuArgAspValThrAlaGlnLeuGln 356
 QY 799 - - - - -GAAATTAAGATCAGGAATAGAGAAATGAGATA 834
 DB 357 SerTyrLysSerSerThrLeuLysGlnIleGlnLysPheLysGlnLeuThrLeu 376
 QY 835 GAA- - - - -CTGGATGAAAGCAACCAACACTTGGAACAGAG- - - - - 870
 DB 377 GlnGlnLysValAlaMetAlaGlnLysSerValGlnAspValGlnGlnGlnIleLeuThr 396
 QY 871 - - - - -CAGCAGAGGAGGCC 885
 DB 397 AlaGlnSerThrAsnGlnGlnThrAlaArgMetValGlnAspLeuGlnAsnArgSerThr 416
 QY 886 CTGGCCAGAGAGAGTGGCTGAGACTACA- - - - -GAACTGCTGGCGGAATCT 933
 DB 417 LeuLysGlnGlnGlnIleLysGlnIleThrSerSerPheLeuGlnLysIleThrAspLeu 436
 QY 934 GAGCACAACCTGCACCTTCACCAAGATCTGAATA- - -GCTCACTGACTCAAGAAAAAGG 990
 DB 437 LysAsnGlnLeuArgGlnGlnAspGlnAspPheArgLysGlnLeuGlnLysGlnLys 456
 QY 991 TATACATATGATTAATTTGGGAAGTTACAGAGAAAGAAATGAAGATTGAGGAACAGTGT 1050
 DB 457 ArgThrAlaGlnLysGlnAsnValMetThrGlnLeuThrMetGlnIleAsnLysTrpArg 476
 QY 1051 GTCCAGCATGGAGAGTACATGAGACAGATGAAGCAAGGCTTAAGAGCTGAT- - - - - 1104
 DB 477 LeuLeuTyrAspGlnLeuTyrGlnLysThrLysProPheGlnGlnGlnLeuAspAlaPhe 496
 QY 1105 - - - - -AAGCAGCCAGGCGCACAGCCAGCAGCTG- - - - - 1134
 DB 497 GluAlaGlnLysGlnAlaLeuLeuAsnGlnLysGln- - -AlaThrGlnGlnGlnLeuAsn 515
 QY 1135 - - - - -GTGACGCTCTCAGCAAGAGAAC- - - - - 1158
 DB 516 LysIleLeuArgAspSerTyrAlaGlnLeuLeuGlnLysHisGlnLeuLysGlnLysIleLys 535
 QY 1159 - - - - -CAGCTTCTCTGGAGAGGAGAGGCTGTGCGAAGAGGTGAGACGGCTCGG 1209

```

Db      536 HisValValLysLeuLysAspGluAsnSerGlnLeuLysSerGlnValSerLysLeuArg 555
QY      1210 ACCAGGTACCAGCATCCACATCTGAT 1239
Db      556 SerGlnLeuValLysArgLysGlnAsnGlu 565

RESULT 5
US-09-310-187A-1
; Sequence 1, Application US/09310187A
; Patent No. 6358751
; GENERAL INFORMATION:
; APPLICANT: Benichou, Gilles
; APPLICANT: Fedoseyeva, Eugenia
; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
; FILE OF INVENTION: Draft Rejection
; FILE REFERENCE: UCSF-090
; CURRENT APPLICATION NUMBER: US/09/310,187A
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-310-187A-1

Alignment Scores:
Pred. No.:      3,94e-13      Length:      1939
Score:          230.50      Matches:      109
Percent Similarity: 42.95%      Conservative: 89
Best Local Similarity: 23.64%      Mismatches: 178
Query Match:      8.44%      Indels:      85
DB:              4          Gaps:      17

US-09-502-945-1 (1-1552) x US-09-310-187A-1 (1-1939)
QY      25 AAACCTACTATGAGCAAAAGTGT----- 48
Db      993 LysLeuThrLysGlnLysLysAlaLeuGlnAlaHisGlnAlaLeuAspLysLeu 1012
QY      49 GAATTTGAGCAATCCCAATGAGTTTGTGAGACAGACTTACTGTAATTCAGAGACT 108
Db      1013 GlnValGlnLysLysLysValAsnSerLeuSerLysSerLysValLysLeuGlnGln 1032
QY      109 TGTGAAGATCTTAAGAGCAACTA---AAGCATAAAGAAATTTCTGCTGCTAATATCT 165
Db      1033 ValAspAspLeuLysLeuSerLeuGlnGlnLysLysValArgMetAspLeuGlnArg 1052
QY      166 TGTAAACCGTGTGGTCTTTTGTGAATGTGCTCAGCATGAGCTGTTCTTCCCAA 225
Db      1053 AlaLysArgLysLeuGlnLysAspLeuLysLeuThrGln---GlnSerIleMetAspLeu 1071
QY      226 ACCCATCAATTAATTCATATGACAGACCATCGAAGAACTGTTAAAGAAAGATGACTTG 285
Db      1072 -----GlnAsnAspLysLeuGlnLysLeuGlnLysLysLysGlnPheAspIle 1089
QY      286 ATGCTCGACTACTTCCGTAAAGAGACAGCTTGCGATACCGCAAGAAAGAGCAAGT 345
Db      1090 AsnGlnGlnAsnSerLysIleGlnAspGlnGlnAlaLeuAlaLeuGlnLysLys 1109
QY      346 GCTTAATGAACAGGTGAACAGATTTTGCAATATCTGAGAGAACCAATTTTGAAGAAAC 405
Db      1110 LeuLysGlnAsnGlnAlaArgIleGlnLysLeuGlnLysLeuGlnLysArgThr 1129
QY      406 AAGCTTTAATCCAGTGTGACGAGTGGAG-----AAGAGCTGGAGAGCGAG 453
Db      1130 ---AlaArgAlaLysValAluLysLeuArgSerAspLeuSerArgLysLeuGlnIle 1148
QY      454 GCGAGACGAGCTTGA----- 468
Db      1149 SerGlnArgLeuGlnLysAlaGlnLysAlaThrSerValGlnIleGlnMetAsnLysLys 1168

```

```

QY      469 AAGAACTTCATCTCAGCAAGAAAGGCCATTCAGAAAGACATGATGAAAAAGAA 538
Db      1169 ArgGlnAlaGlnLysLeuGlnLysMetArgArgAspLeuGlnLysLeuGlnHisGln 1188
QY      529 ATAACGAAAGAA-----AGGAGTACATGGGATCAAAAGATGTGTGCTCAGAA 582
Db      1189 AlaThrAlaAlaLeuArgLysLysHisAlaAspSerValAlaGlnLeuGlnGln 1208
QY      583 ATTGCCCACTGGAGGCCAGTGGAAAGGTTACAAAGGAAAGAAATTCAGCTAATTAAT 642
Db      1209 IleAspAsnLeuGlnArgValLysGlnLysLeuGlnLysLysSerGlnPheLysLeu 1228
QY      643 CAACGTGAGGAAATTAACAC-----CACTGGCTTCTCGGAAATGATGTCTACA 693
Db      1229 GlnLeuAspAspValThrSerAsnMetGlnIleIleLysAlaLysAlaAsnLeuGln 1248
QY      694 AAGGTGTGTGAGAAATGCCCTATCAGCTGAATTAACCAACTGGAGAGATGAGCGCA 753
Db      1249 LysValSerArgThrLeuGlnAspGlnAlaAsnGlnLysArgValLysLeuGlnAla 1268
QY      754 GAAAGAGCAGCAGAGAGTTC-----AGAGCAAAACTAAC----- 789
Db      1269 GlnArgSerLeuAsnAspPheThrThrGlnArgAlaLysLeuGlnThrGlnAsnGln 1288
QY      790 -----AGGATCTTGAATTAAGATCAGAAATGAGAAATG----- 828
Db      1289 LeuAlaArgLysLeuGlnLysGlnLysAlaLeuIleSerGlnLeuThrArgGlyLysLeu 1308
QY      829 -----AGATACAGCTGAGTGAACCAACCAACTGTGACAGAG-----CAGCAGAG 879
Db      1309 SerTyrThrGlnIleMetGlnAspLeuLysArgGlnLeuGlnLysLysAlaLys 1328
QY      880 GCAGCCCTGGCC----- 900
Db      1329 AsnAlaLeuAlaHisAlaLeuGlnSerAlaArgHisAspLysAspLeuLeuArgGln 1348
QY      901 TGCCTGAGACTAACAGAACTGCTGGCGAATCTGAGACCAACTGCACCTCACCAGATCT 960
Db      1349 TyrGlnGlnLysThrGlnAlaLysAlaGlnLeuGlnArgValLeuSerLysAlaAsnSer 1368
QY      961 GAATAGCTCACTCACTCACTCAAGAAAGGTTATCATATGATTAATGGAAAGTTACAG 1020
Db      1369 GlnValAlaGln-----TrpArgThrLysTyrGlnThrAspAlaIle----- 1382
QY      1021 AGAAGAAATGAAAGATTTGAGAGCAAGTGTGTCAGCATGGAGAGTACATGAGAGCATG 1080
Db      1383 GlnArgThrGlnLysLeuGlnLys-----AlaLysLysLysLeu 1395
QY      1081 AAGCAAAAGCTTAAGCACTGGATTAAGCACAGCCAGCCACAGCCACAGCTGTGTCAG 1140
Db      1396 AlaGlnArgLeuGlnAspAlaGlnLysAlaValAlaGlnAlaValAsnAlaLysSerSer 1415
QY      1141 CTCCTCAGACAGAGCAAGCAAGCTTCTCCTGAGAGAGAGAGCCTGTGCGAAGAGTGGAG 1200
Db      1416 LeuGlnLysThrLysHisArgLeuGlnAsnGlnIleLysAspLeuMetValAspValGln 1435
QY      1201 CCG 1203
Db      1436 Arg 1436

RESULT 6
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; FILE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN

```

STREET: 1601 MARKET STREET, SUITE 720
 CITY: PHILADELPHIA
 STATE: PA
 COUNTRY: USA
 ZIP: 19103-2307
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/353,700
 FILING DATE: 09-DEC-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: REED, JANET E.
 REGISTRATION NUMBER: 36,252
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (215) 563-4100
 TELEFAX: (215) 563-4044
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3248 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: HUMAN
 US-08-353-700-1

Alignment Scores:
 Pred. No.: 9.39e-13 Length: 3248
 Score: 227.50 Matches: 97
 Percent Similarity: 37.55% Conservative: 93
 Best Local Similarity: 19.17% Mismatches: 175
 Query Match: 8.33% Indels: 141
 Gaps: 15

US-09-502-945-1 (1-1552) x US-08-353-700-1 (1-3248)

QY 1 CTTCTGATGATCCGAGAACCTAAACTTACTTATGAGAAAGTGAATGAGGAA 60
 Db : : : : : ||||| ||| : : : : :
 75 IIEcysgluSerleuGluLysThrLysGlnLysIleSerHisGluLeuGlnValLysLeu 94
 QY 61 TCCCAATGAATTTTGGAGAACGACTTACCTGAATATCAGAGAACTTGAAGATCTT 120
 Db : : : : : ||||| ||| : : : : :
 95 SerGlnValAsnPhenGlnGluGlnLeuAsnSerGlyLysGlnIleGlnLysLeu 114
 QY 121 AAAAGCAACTAAAGCAATAAAGATTTCTTGCGCTGCTAATCTGTAAACCGTGTGCT 180
 Db : : : : : ||||| ||| : : : : :
 115 GluGlnGluLeuLys----- 119
 QY 181 GGTCTTTGTTGAATGTCTCAGCAGTAAAGCTGTTCTTCCAAACCATTAATGTT 240
 Db : : : : : ||||| ||| : : : : :
 120 -----ArgCysLys----- 122
 QY 241 CATATGACACCATGAAAGACTGGTTAAAGAAAGATGACTTGATGCTGCAGTACTT 300
 Db : : : : : ||||| ||| : : : : :
 123 -----SerGluLeuGlnValArgSerGlnGlnIleAlaIleSerAlaAspVal 137
 QY 301 TCCGTAAG-----AGCAGCTTGGCGAGATACGACG 330
 Db : : : : : ||||| ||| : : : : :
 138 SerLeuAsnProCysAsnThrProGlnLysIlePheThrProLeuThrProSerGln 157
 QY 331 CAAGAAGAACGAGCTTATGACAGAGTGAACAAGTTTG-----CAATATCTGAG 384
 Db : : : : : ||||| ||| : : : : :
 158 TyrTyrSerGlySerLysTyrGlnAspLeuLysGlnLysTyrAsnLysGlnValGlnGln 177
 QY 385 GAAGCCAAATTTTGA-----AAAACCAAGGCTTTA 414
 Db : : : : : ||||| ||| : : : : :

Db 178 ArgLysArgLeuGlnValAlaGlnValLysAlaLeuGlnAlaLysLysAlaSerGlnThrLeu 197
 QY 415 ATCCAGTGTGACCACTTGAGAGAGAGCTGGAGAG----- 450
 Db : : : : : ||||| ||| : : : : :
 198 ProGlnAlaThrMetAsnHisArgAspIleAlaArgHisGlnAlaSerSerValPhe 217
 QY 451 -----CAGCGGAGAGCACTTGAAAGAACCTGCATCTGCAGCAAGAAAGGCCATT 504
 Db : : : : : ||||| ||| : : : : :
 218 SerTrpGlnGlnGlnLysThrProSerHisLeuSerSerAsnSerGlnArgThrProIle 237
 QY 505 GAGAAAGACATGATGAAAGAAATACGAAGAAAGAGAGTCAATGGAATCAAGATG 564
 Db : : : : : ||||| ||| : : : : :
 238 ArgArgAspPheSerAlaSerTyrPheSerGlyGlnLeuGlnValThrProSerArgSer 257
 QY 565 TTGATC----- 570
 Db : : : : : ||||| ||| : : : : :
 258 ThrLeuGlnIleGlyLysArgAspAlaAsnSerSerPheGlyAsnSerSerPro 277
 QY 571 -----TTGCTCTCAGAAATATGCCCCAAGCTG 594
 Db : : : : : ||||| ||| : : : : :
 278 HisLeuLeuAspGlnLeuLysAlaGlnAsnGlnGlnLeuAlaGlnLysIleAsnGlnLeu 297
 QY 595 GAGGCCAGGTGAGAAAGCTTACAAAGGAAAGATTCAGCTATTATCAACTGAGGAA 654
 Db : : : : : ||||| ||| : : : : :
 298 GluLeuArgLeuGlnGlnLysHisGlnLysGlnMetLysGlnValAsnLysPheGlnGln 317
 QY 655 ATTCAAGCCACGCTGGCTTCGCGGAATGATGCTC-----ACAAAGTGTGGA 705
 Db : : : : : ||||| ||| : : : : :
 318 LeuGlnLeuGlnLeuGlnLysAlaLysAlaGlnLeuIleGlnLysGlnLysValLeuAsn 337
 QY 706 GAAATGCGCTATCAGCTGATGAAATGAAACCAATGAGAGATGAGAGCAAGAAAGAGAC 765
 Db : : : : : ||||| ||| : : : : :
 338 LysCysArgAspGlnLeuValAlaArgThrThrAlaGlnThrAspGlnAlaSerThrLysTyr 357
 QY 766 AGAGATTGAGAGCAAAAGCTAACAG-----GATCTGAAATTAAGATCAGGAA 816
 Db : : : : : ||||| ||| : : : : :
 358 ThrAlaLeuGlnGlnLysLeuLysLeuThrGlnAspLeuSerCysGlnArgGlnAsn 377
 QY 817 ATAGAGAAATGAGAAATGAACTGAT-----GAAAGCAA 852
 Db : : : : : ||||| ||| : : : : :
 378 AlaGlnSerAlaArgCysSerLeuGlnLysIleLysGlnLysGlnLysGlnPheGln 397
 QY 853 CAACACTTGGAAACAGACAGACAGAACGACCGCTGGCCAGAGAGAGTGCCTG----- 906
 Db : : : : : ||||| ||| : : : : :
 398 GluGlnLeuSerArgGlnGlnArgSerPheGlnThrLeuAspGlnGlnLysIleGlnMet 417
 QY 907 -----AGCTAACAGAACTGCTGGCGAATCTGAGCACCAATGCACTCACCAGATCT 960
 Db : : : : : ||||| ||| : : : : :
 418 LysAlaArgLeuThrGlnGlnLysGlnGlnIleAlaLysAsnMetHisAsnValLeuGlnAla 437
 QY 961 GAAATGAGCTCAAGCTCAGTCAAGAAAGAAAGTATACATATGATTAATTTGGAAAGTTACAG 1020
 Db : : : : : ||||| ||| : : : : :
 438 GluLeuAspLysLeuThrSerValLysGln-----GlnLeuGln 450
 QY 1021 AGAAGAAATGAAGATTGAG-----GAACAGTGTGTCCACATGGG 1062
 Db : : : : : ||||| ||| : : : : :
 451 AsnAsnLeuGlnGlnLysPheLysGlnLysLeuCysArgAlaGlnGlnAlaPheGlnAlaSer 470
 QY 1063 AGACTTACTGAGAG-----ATGAAGCAAAAGGCTTAAGGCAG 1098
 Db : : : : : ||||| ||| : : : : :
 471 GlnIleLysGlnLysAsnGlnLeuAlaArgSerMetGlnGlnMetLysGlnAsnAsnLeu 490
 QY 1099 CTGATTAAGACACAGCCAGCCAGCCAGAGCTGTGTCAGCTC-----CTCAGC 1149
 Db : : : : : ||||| ||| : : : : :
 491 LeuLysSerHisSerGlnGlnLysAlaArgLysHisLeuGlnValAlaGlnLeuLys 510
 QY 1150 AAGCAGAACCCAGCTTCTCTCCTGAGAGAGAGAGCTGCGAAGAGTGCACCGGCTCGG 1209
 Db : : : : : ||||| ||| : : : : :
 511 AsnIleLysGlnCysLeuAsnGlnSerGlnAsnPheAlaGlnGlnMetLysAlaLysAsn 530
 QY 1210 ACCCAGTTACCCAGCAGT 1227
 Db : : : : : ||||| ||| : : : : :
 531 ThrSerGlnGlnThrMet 536

```

RESULT 7
PCT-US95-16216-1
: Sequence 1, Application PC/TUS9516216
: GENERAL INFORMATION:
: APPLICANT: Yen, Timothy J.
: APPLICANT: Raltner, Jerome B.
: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ. ID NO. 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Alignment Scores:
Pred. No.: 9, 39e-13 Length: 3248
Score: 227.50 Matches: 97
Percent Similarity: 37.55% Conservative: 93
Best Local Similarity: 19.17% Mismatches: 175
Query Match: 8, 33% Indels: 141
DB: 5 Gaps: 15

US-09-502-945-1 (1-1552) x PCT-US95-16216-1 (1-3248)
QY 1 CTTCTGATGATCCGAGAGCTAAACTTACTTATGAGGAAAGTGTGAATTGAGAA 60
DB 75 Tlecygluserleuylulsthrlysglnlylleserhisluleuglnvalylsgln 94
QY 61 TCCCAATGAAGTTTGTGAGGAACGACTTAGCTGAATATCAGAGAACTGTGAAGTCTT 120
DB 95 SerGlnValAsnphelnglnlueuAsnserGlyLyslysglnlleglulstleu 114
QY 121 AAGAGCACTAAGCATTAAGAAATTTCTTGCGTGCATTAATCTGTAAACGCTGTGGT 180
DB 115 GluInglnluleuys----- 119
QY 181 GGTCTTTGTTGAATGTGCTCAGATGAAGCTGTCTTCCCAACCATACTAATGTT 240
DB 120 -----Argcyslyls----- 122
QY 241 CATATGACAGACATCGAAAGACTGGTTAAAGAAAGAGATGACTGTGTCGCACATGTT 300

```

```

DB 123 -----sergluleuglnArgserGlnGlnlaalaGlnserAlaAspval 137
QY 301 TCCGTAAG-----AGCAGCTTGGCAGATACGCG 330
DB 138 SerleuAsnProCysAsnThrProGlnLysllePheThrProleuThrProSerGln 157
QY 331 CAAGAGAAAGCAAGCTCTTATGAACAGGTGAACAGATTGTTG-----CAATATGTCAG 384
DB 158 TyrTyrSerGlySerLysTyrGlnAspLeuLysGlnLysTyrAsnLysGlnValGln 177
QY 385 GAAGCCAAATTGCA-----AAAGCCAAAGCTTTA 414
DB 178 ArgLysArgLeuGlnlualGlnValLysAlaLeuGlnAlaLysLysAlaSerGlnThrLeu 197
QY 415 ATCCAGTGTGACCAAGTTGAGGAGGAGCTGAGAG----- 450
DB 198 ProGlnAlaThrMetAsnHisArgAspIleAlaArgHisGlnAlaSerSerValPhe 217
QY 451 -----CAGCGGAGCGACTTGAAGAAAGAACTTCATCTCAGCAGAGAAAGGCCATT 504
DB 218 SerTrrGlnGlnGlnLulsthrProSerHisLeuSerSerAsnSerGlnArgThrProIle 237
QY 505 GAGAAAGACATGATGAAGAAAGAAATACGAAGAAAGAGAGATCATGGATCAAGATG 564
DB 238 ArgArgAspPheSerAlaSerTyrPheSerGlyLuleuGlnValThrProSerArgSer 257
QY 565 TTGATC----- 570
DB 258 ThrLeuGlnlleGlyLysArgAspAlaAsnSerPhePheGlyLysnSerSerPro 277
QY 571 -----TTGTCAGATAATGGCCCACTG 594
DB 278 HisLeuLeuAspGlnleuLysAlaGlnAsnGlnGlnleuArgAsnLysIleAsnGlnleu 297
QY 595 GAGGCCAGGTGAGAAAGGTACAAAGAGAAAGATTTCACCTTATATCAGCAGAGAA 654
DB 298 GluLeuArgLeuGlnlGlnLysGlnLysGlnMetLysGlyGlnValAsnLysPheGlnGln 317
QY 655 ATTCAAAGCCAGCTGCTCTCGGAAATGAGATGC-----ACAAAGCTGTGGA 705
DB 318 LeuGlnleuGlnleuGlnLysAlaLysValGlnleuIleGlnLysGlnLysValIleuAsn 337
QY 706 GAAATGCGTATCAGCTGATGAATAAACCAACATGAGAGAGATGAGGACAGAAAGAGCAC 765
DB 338 LysCysArgAspGlnleuValArgThrThrAlaGlnLtyArgAspLlnAlaSerThrLysTyr 357
QY 766 AGAGAGTTAGAGCAAAACTAAGCAG-----GATCTTGAATTTAAAGTCAGGAA 816
DB 358 ThrAlaLeuGlnlGlnLysLeuLysLysLeuThrGlnAspLeuSerCysGlnArgGlnAsn 377
QY 817 ATAGAGAAATTTGAGATTAAGACTGAT-----GAAGACAA 852
DB 378 AlaGlnSerAlaArgCysSerleuGlnGlnLyslleLysGlnLysGlnLysGlnPheGln 397
QY 853 CAACACTTGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 906
DB 398 GluGlnleuSerArgGlnGlnArgSerPheGlnThrLeuAspGlnLysIleGlnMet 417
QY 907 -----AGACTACAGACACTGCTGGCGAATCTGAGCACCACACTGACCTACACAGATCT 960
DB 418 LysAlaArgLeuThrGlnlGlnleuGlnGlnAlaLysAsnMetHisAsnValleuGlnAla 437
QY 961 GAAATGTCTCACTCAGTCAGCAAGAAAGAGTATACATATGATTAATGGGAAAGTACAG 1020
DB 438 GluLeuAspLysLeuThrSerValLysGln-----GlnleuGln 450
QY 1021 AGAAGAAATGAAGAAATTTGAG-----GAACAGTGTGTCAGCATGAG 1062
DB 451 AsnAsnleuGlnGlnLuePheLysGlnLysLeuCysArgAlaGlnGlnAlaPheGlnAlaSer 470
QY 1063 AGAGTACATGAGACG-----ATGAGCAAAAGGCTAAGGCAG 1098

```

Alignment Scores:		
Pred. No.:	1,39e-12	length: 1368
Score:	224.00	Matches: 112
Percent Similarity:	38.36%	Conservative: 94
Best Local Similarity:	20.86%	Mismatches: 163
Query Match:	6.20%	Indels: 148
Db:	4	Gaps: 17

QY	307	AGGAGACAGCTGGCAGATPACGACGAAGAAGCAAGCAAGTCATTATGACAGGTGAAACAA	366
Db	995	ArgSerSerValLysGlu-----LysThrGluThrIleSerPheLysGln	1010
QY	367	GTTTTGCAAAATATCTGAGAGAGCAATTTTGAAAAAACAGGCTTTAATCCAGTGTAC	426
Db	1011	GluLeu-----LysAspIleAsnLysTyrAsnSerAlaLeuValAspArgGlu	1027
QY	427	CAGTTGAGAGAGAGAGCTGGAGAGGACG-----	465
Db	1028	GluSerArgValLeuIleLysLysGlnGluValAspIleLeuAspLeuLysGluThrLeu	1047
QY	466	GAAGAAAGAACTTCGATCTGCACGAAGAAAGGAGCCATTGAGAAAGACATGATGAAGAG	525
Db	1048	ArgLeuArgIleLeuSerGluAsp-----IleGluArgAspMetLeuLysGlu	1065
QY	526	GAATTAACGAAGAAAGAGAGTACATGGATCAAGAGATGTGATCTGTCCAGATATT	585
Db	1064	AspLeuAlaHisAlaThrGlnGluLeuAsnMetLeuThrGluAlaSerLysLysHisSer	1083
QY	586	GCCCACTGGAGGGCCGAGTGGAAAGAGTTACAAAGAAAGATTTCCAGCTATTAAACA	645
Db	1084	GlyLeuLeuGlnSerAlaGlnGluGluLeuThrLysLysGluAla-----Leu	1093
QY	646	CTGGAGGAATTCAAAGCCAGTCCGCTTCCGGAAATGATGTCAAGAGTGTGTGA	705
Db	1100	IleGlnGlnLeuGlnHisLysLeuAsnGlnLysLysGluGluValGluGlnLysAsn	1119
QY	706	GAATGGCGCTNACGCTGAATTAACCAACATGGAGAAAGATGAGGCA-----	753
Db	1120	GluTyrAsnPheLysMetLarGlnLeuGluHisValMetAspSerAlaIleGluAspPro	1133
QY	753	-----	753
Db	1140	GlnSerProLysThrProPheHisPheGlnThrHisLeuAlaLysLeuLeuGluThrGln	1159
QY	754	GAAGAAGCACACAGAGCTTCAAGCA-----	783
Db	1160	GluGlnGluIleGluAspArgLysArgAlaSerLysThrSerLeuGlnHisLeuValThrLys	1179
QY	784	ACTGAACGGAGCTTGTCAAAATTAAGATTCAGAAATA-----	819
Db	1180	LeuAsnGluAspArgGluValLysAsnAlaGluIleLeuArgMetLysGlnLeuArg	1199
QY	820	-----GAGAAATTGAGATAGAA-----	837
Db	1200	GluMetGluAsnLeuArgLeuGlnSerGlnGlnLeuIleGluLysAsnTyrPheLeuGln	1219
QY	838	-----CTGGATGAAGACAA-----	852
Db	1220	GlyGlnLeuAspSerIleLysArgGlnLysGluAsnSerAspGlnAsnHisProAspAsn	1233
QY	853	CAACACTTGAACAGACAGCAGCAG-----AAGCAGCCCTGGCCAGA-----	894
Db	1240	GlnGlnLeuLysAsnGlnGlnGlnGlnSerIleLysGluArgLeuAlaLysSerLysIle	1259
QY	895	---GAGAGTGGCCGTAGACATCAACAGAACATCGTGGCGGATCTGCACCAACATCGACCTC	951
Db	1260	ValIleGlnMetLeuLysMetLysAlaAspLeuGluGluValGlnSerAlaLeuTyrAsn	1279
Db	1300	LysAlaPheGlnGluLysGlnGlnLeuAspSerLysLeuGlnGluMetLysTyrGluArg	1319
QY	1012	AAGTTACAGAGAAATTAAGATTGAGAGAACAGTGTCTCCACAGTGGAGAGTACAT	1071
Db	1280	LysGluMetGluLysLeuArgMetThrAspGluValGluArgThrGlnThrLeuGlnSer	1299
QY	1072	GAGACGATGAGCAAGCAAGGCTTAAGGACGTGGATTAACACACAGCCAGCCACAGCCAGG	1133
Db	1320	GluArgThrSerGlnGlnMetGluMetLeuArgLysGlnValGluLysLeuAlaGluGlu	1339
QY	1132	CTGGTCAGCTCTCCACAGCAAGCAAC-----CAGCTT	1164

Db 1340 AsnGlyLysLeuValGlnHisGlnAsnLeuHisGlnLysIleGlnTyraValAlaArgLeu 1359
QY 1165 CTCTGTGAGAGCGAGAGCCTGTGCGAAGAGAGTGGCGGCTGCGACCCAG 1215
Db 1360 LysLysGlnAsnValAlaArgLeuAlaGlnGluThrGlnLysLeuArgAlaGln 1376
RESULT 9
US-09-723-262-2
; Sequence 2, Application US/09723262
; Patent No. 6379912
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6379912el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,262
; CURRENT FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: US 09/572,191
; PRIOR FILING DATE: 2000-05-17
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRF
; ORGANISM: Human
US-09-723-262-2

Alignment Scores:
Pred. No.: 1,39e-12 Length: 1388
Score: 224.00 Matches: 112
Percent Similarity: 38.36% Conservative: 94
Best Local Similarity: 20.86% Mismatches: 183
Query Match: 8.20% Indels: 148
Gaps: 17

US-09-502-945-1 (1-1552) x US-09-723-262-2 (1-1388)

QY 1 CTCTGTGAGAGCGAGAGCCTGTGCGAAGAGTGGCGGCTGCGACCCAG 60
Db 856 LeuLeuGlnLysLysValAlaCysLeuGlnAspSerTyraAspAsnLeuGlnIleMetLys 875
QY 61 TCCCAATGGAAGTTTGTGAGGAGCACTTACGTGAATATCAGAGCACTGTGGAAGTCTT 120
Db 876 PheGlnIleAspGlnLysSerArgAsnLeuGlnAsnPheLysLysGlnAsnGlnIleThrLeu 895
QY 121 AAAGACCACTAAAGCATTAAGAAATTTCTTGTGCTGCTAATCTGTAAACCGTGTGGT 180
Db 896 LysSerAspLeuAsnAsnLeuMetGlnLeuLeuGlnAlaGlnLys--GlnArgAsnAsn 914
QY 181 GGCTCTTGTGTTG----- 192
Db 915 LysLeuSerLeuGlnPheGlnGlnAspLysGlnAsnSerSerLysGlnIleLeuLysVal 934
QY 193 -----AAATGTGCTCAGCATGGAAGCT 213
Db 935 LeuGlnAlaValAlaArgGlnGlnLysGlnLysGlnIleThrAlaLysCysGlnGlnIleMetAla 954
QY 214 GTTCTTTCCCAAAACCATATCAATGTT-----CATATGCAGACCATCGAAAGACGTGGT 267
Db 955 LysValGlnLysLeuGlnLysSerLeuLeuAlaThrGlnLysValIleSerSerLeuGln 974
QY 268 AAAGAAGAAGAT-----GACTGTGATGTCGACACAGTTTCCGTA 306
Db 975 LysSerArgAspSerAspLysLysValValAlaAspLeuMetAsnGlnIleGlnIleLeu 994
QY 307 AGGAGCAGCTTGGCAGATACGAGCAAAAGAGCAAGTCTTATGACAGAGTGAACAA 366
Db 995 ArgSerSerValCysGln-----LysThrGlnThrIleAspThrLeuLysGln 1010
QY 367 GTTTTGCAAATATCTGAGAGAGCCAAATTTTGAACCAAGGCTTTAATCCAGTGTGAC 426

Db 1011 GlnLeu-----LysAspIleAsnCysLysTyraAsnSerAlaLeuValAspArgGln 1027
QY 427 CAGTTGAGAGAGAGCGAGAGAGCAG-----GGGAGCGACCTT 465
Db 1028 GlnSerArgValLeuIleLysLysGlnGlnValAlaAspIleLeuAspLeuLysGlnIleThrLeu 1047
QY 466 GAAAAAGACCTTCATCTCAGCAAGAGAAAAAGGCCATTGAGAAAGCATGATGAAAAAG 525
Db 1048 ArgLeuArgIleLeuSerGlnAsp-----IleGlnArgAspMetLeuCysGln 1063
QY 526 GAATATACGAAAGAGAGGATGATCGGATCAAGATGTTGATCTTGTCTCGAATAT 565
Db 1064 AspLeuAlaHisAlaThrGlnGlnLeuAsnMetLeuThrGlnAlaSerLysLysHisSer 1083
QY 586 GCCCACTGGAGCCCGAGTGGTGTGATACAAAGGAAAGATTTCAGCTATTATATCA 645
Db 1084 GlyLeuLeuGlnSerAlaGlnGlnGlnIleLeuThrLysGlnAla-----Leu 1099
QY 646 CTGAGAGAAATTCAGAACCCAGTGGCTTCTCGGAAATGATGTACAAAGGTGTGGA 705
Db 1100 IleGlnGlnLeuGlnHisLysLysLeuAsnGlnLysLysGlnGlnValGlnIleLysAsn 1119
QY 706 GAATGCGCTATCAGCTGAATTAACCAACATGAGAGATGAGCA----- 753
Db 1120 GlnTyraAspPheLysMetArgGlnLeuGlnHisValMetAspSerAlaAlaGlnAspPro 1139
QY 753 ----- 753
Db 1140 GlnSerProLysThrProPheHisPheGlnThrHisLeuAlaLysLeuGlnIleThrGln 1159
QY 754 GAAAGAGAGCAGAGAGTTCAGAGCA-----AAA 783
Db 1160 GlnGlnGlnIleGlnAspArgLysArgAlaSerLysThrSerLeuGlnHisLeuValThrLys 1179
QY 784 ACTTAACAGGATCTTGAATTAAGATCAGAGAAATA----- 819
Db 1180 LeuAsnGlnAspArgGlnValLysAsnAlaGlnIleLeuArgMetLysGlnIleLeuArg 1199
QY 820 -----GAGAAATTTGAGAAATGAA----- 837
Db 1200 GlnMetGlnAsnLeuArgLeuGlnSerGlnGlnLeuIleGlnLysAsnThrPheLeuGln 1219
QY 838 -----CTGGATGAAGCAAA----- 852
Db 1220 GlyGlnLeuAspArgLysArgGlnLysGlnLysSerAspGlnAsnHisProAspAsn 1239
QY 853 CAACACTTGGACACAGAGCAG-----AAGCAGCCCTGGCCAGA----- 894
Db 1240 GlnGlnLeuLysAsnGlnGlnGlnIleGlnIleLysGlnArgLeuAlaLysSerLysIle 1259
QY 895 -----GAGGAGTCCCTGAGACTTAACAGAACTGCTGGGAGCACTGACACCACTGCACCTC 951
Db 1260 ValGlnGlnMetLeuLysMetLysAlaAspLeuGlnGlnValGlnSerAlaLeuTyraAsn 1279
QY 952 ACCAGATCTGAATATAGCTCAACTCACTCAAGAAAAAGGTATATCATATGATAAATGGGA 1011
Db 1280 LysGlnMetGlnCysLysLeuArgMetThrAspArgLysValGlnArgGlnThrLeuGlnSer 1299
QY 1012 AAGTTACAGAGAAAGAAATGAGAAATTTGGAGAAACATGTGTCTCAGCATGGGAGATACAT 1071
Db 1300 LysAlaPheGlnGlnLysGlnGlnLeuArgSerLysLeuGlnIleMetTyraGlnIleArg 1319
QY 1072 GAGACATGAGCAAGAGGCTTAAGGCGACTGGATTAACACAGCCAGCCAGCCAGCCAG 1131
Db 1320 GlnArgThrSerGlnIleMetGlnMetLeuArgLysGlnValGlnCysLeuAlaGlnGln 1339
QY 1132 CTGGTCCAGCTCTCTCAGCAAGAGAAC-----CAGCTT 1164
Db 1340 AsnGlyLysLeuValGlnHisGlnAsnLeuHisGlnLysIleGlnTyraValAlaArgLeu 1359
QY 1165 CTCTGTGAGAGCGAGAGCCTGTGCGAAGAGTGGCGGCTGCGACCCAG 1215

Db 1360 LysLysGluAsnValArgLeuAlaGluThrGluLysLeuArgAlaGlu 1376

RESULT 10
US-09-723-219-2
; Sequence 2, Application US/09723219
; Patent No. 6391613
; GENERAL INFORMATION:
; APPLICANT: Berard, Christophe
; APPLICANT: Sakowicz, Roman
; APPLICANT: Wood, Kenneth
; TITLE OF INVENTION: No. 6391613el motor proteins and methods for
; FILE REFERENCE: 1017
; CURRENT APPLICATION NUMBER: US/09/723,219
; PRIOR APPLICATION NUMBER: US 09/572,191
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1388
; TYPE: PRN
; ORGANISM: Human
US-09-723-219-2

Alignment Scores:
Pred. No.: 1.39e-12 Length: 1388
Score: 224.00 Matches: 112
Percent Similarity: 38.36% Conservative: 94
Best Local Similarity: 20.86% Mismatches: 183
Query Match: 8.20% Indels: 148
Gaps: 17

US-09-502-945-1 (1-1552) x US-09-723-219-2 (1-1388)

QY 1 CTTCTGATGCATCCGAGAGCTAAACTTACTTATGAGGAAAGGTGTAATTGAGGA 60
Db 856 LeuLeuGlnSerLysAlaCysLeuGlnAspSerTyraSpAsnLeuGlnLileuMetLys 875
QY 61 TCCCAATGAAAGTTTGTGAGAGACACTTACGTGAATATCAGAGAACTGTGAATCTT 120
Db 876 PheGluIleAspGlnLeuSerArgAsnLeuGlnAsnPheLysGluAsnGluThrLeu 895
QY 121 AAAGAGCACTAAAGCAATTAAGAAATTTCTTGCTGCTGAATACTGTATACCGTGTGGT 180
Db 896 LysSerAspLeuAsnAsnLeuMetLileuLeuGlnAlaGluLys---GluArgAsnAsn 914
QY 181 GGTCTTGTGTTG----- 192
Db 915 LysLeuSerLeuGlnPheGlnGluAspLysGluAsnSerLysGluIleLeuLysVal 934
QY 193 -----MAATGTCCTCAGCATTAAGCT 213
Db 935 LeuGlnAlaValArgGlnGlnLysGlnLysGlnThrAlaLysCysGlnGlnGlnMetAla 954
QY 214 GTTCTTCCCAACCACTACTAATGTT-----CATATGCAGACCATGGAAGACTGGTT 267
Db 955 LysValGlnLysLeuGlnGlnSerLeuLeuAlaThrGluLysValIleSerSerLeuGln 974
QY 268 AAAGAAAGAGAT-----GACTTGATGCTGCAGTAAGTTCCGTA 306
Db 975 LysSerArgAspSerAspLysLysValValAlaAspLeuMetAsnGlnIleGlnLeu 994
QY 307 AGGAGACCTTGGCAGATACGACGAAAGAGAGCAAGTCTTATGAACAGGTGAACAA 366
Db 995 ArgSerLeuValCysGln-----LysThrGlnThrIleAspThrLeuLysGln 1010
QY 367 GTTTTGCAAAATATCGAGAGCAATTTTGAAAAAACAAGGCTTTATCCAGTGTAC 426
Db 1011 GluLeu-----LysAspIleAsnCysLysTyraSerAlaLeuValAspArgGlu 1027
QY 427 CAGTTGAGAGAGAGCTGAGAGGACG-----GCGAGCGAGCTT 465
::: ||| ::::::::::||| ||| |||

Db 1028 GluSerArgValIleuLileuLysGlnGlnValAspIleLeuAspLeuLysGlnThrLeu 1047
QY 466 GAAAAAGAACTTCATCTTCAGCAAGAAAGGCCCTTGAAAGACATGTGAAAAAG 525
Db 1048 ArgLeuArgIleLeuSerLysLysP-----IleGluArgAspMetLeuLysGln 1063
QY 526 GAAATTAACGAAGAAAGAGGAGTACATGGATCAAGAGTGTATCTTCACGATATT 585
Db 1064 AspLeuAlaHisAlaThrGlnGlnLeuAsnMetLeuThrGlnAlaSerLysHisSer 1083
QY 586 GCCCACTGGAGGCCAGGTGGAAGGTTTACAAAGAAAGATTTCAGCTATTATCA 645
Db 1084 GlyLeuGlnSerAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1099
QY 646 CTGGAGGAAATTCAAAGCCAGCTGCTTCGCGAAATGATGTACAAAGGTGTGTGA 705
Db 1100 IleGlnGlnLeuGlnHisLysLeuAsnGlnLysLysGlnGlnValGlnGlnLysLysAsn 1119
QY 706 GAAATGCGCTATCAGCTGATTAACCAACATGAGAGAGTATGAGCA----- 753
Db 1120 GluTyraAspPheLysMetArgGlnLeuGlnHisValMetAspSerAlaAlaGlnAspPro 1139
QY 753 ----- 753
Db 1140 GlnSerProLysThrProProHisPheGlnThrHisLeuAlaLysLeuGlnThrGln 1159
QY 754 GAAAAAGAACACAGAGAGTTCAGACA-----AAA 783
Db 1160 GlnGlnGlnIleGlnAspLysArgAlaSerLysThrSerLeuGlnHisLeuValThrLys 1179
QY 784 ACTAAGAGGATCTTGAATTAAGATCAGCAATA----- 819
Db 1180 LeuAsnGlnAspArgGlnValLysAsnAlaGlnIleLeuArgMetLysGlnLeuArg 1199
QY 820 -----GAGAAATGAGAAATAGAA----- 837
Db 1200 GluMetGluAsnLeuArgLeuGlnSerGlnGlnLeuIleGlnLysAsnThrLeuGln 1219
QY 838 -----CTGGATGAAAGCAAA----- 852
Db 1220 GlnGlnLeuAspAspIleLysArgGlnLysGluAsnSerAspGlnAsnHisProAspAsn 1239
QY 853 CAACACTTGGAGACAGAGACAGCAG-----AAGCGAGCCCTGGCCAGA----- 894
Db 1240 GlnGlnLeuLysAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1259
QY 895 ---GAGAGTGGCTGAGACTAACAAGACTGCTGGCGAAATCTGAGCCACCACTGCACCTC 951
Db 1260 ValGlnGlnMetLeuLysMetLysAlaAspLeuGlnGlnValGlnSerAlaLeuTyraAsn 1279
QY 952 ACCAGATCTGAATATAGCTCAACTCAGTCAAGAAAAAGGTATATGATTAATTGGGA 1011
Db 1280 LysGlnMetGlnCysLeuArgMetThrAspGlnValGluArgThrGlnThrLeuGlnSer 1299
QY 1012 AAGTTACAGAGAAATGAAGAAATTTGAGAGCAAGCTGTGCATGAGAGAGTACAT 1071
Db 1300 LysAlaPheGlnGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1319
QY 1072 GAGACGATGAACCAAGGCTAAGGAGCTGTGATTAAGCAACAGCCAGCCACAGCCAGCAG 1131
Db 1320 GluArgThrSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1339
QY 1132 CTGGTGCAGCTCTCTCAGACAGACAGAC-----CAGCTT 1164
Db 1340 AsnGlnLysLeuValGlnHisGlnAsnLeuHisGlnLysIleGlnTyraValValArgLeu 1359
QY 1165 CTCCTGAGAGAGGACAGAGGCTGCGAGAGAGGAGCGGCTGCGAGAGCCAG 1215
Db 1360 LysLysGluAsnValArgLeuAlaGlnIleuThrGlnLysLeuArgAlaGln 1376

RESULT 11
US-08-056-200-94
; Sequence 94, Application US/08056200

[illegible]


```

; REFERENCE/DOCKET NUMBER: 5004-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 955 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-282-845-2

Alignment Scores:
Pred. No.: 2,27e-12 Length: 955
Score: 221.00 Matches: 94
Percent similarity: 41.22% Conservative: 68
Best Local Similarity: 23.92% Mismatches: 165
Query Match: 8.09% Indels: 66
DB: Gaps: 12

US-09-502-945-1 (1-1552) x US-08-282-845-2 (1-955)
OY 1 CTTCTGATGATCCGAGAAAGCTAAACTTATGAGGAAAAGTGAATT----- 54
Db 573 LeuGlnGluAlaLeuGlnSerGluArgThrAlaAlaGluArgSpGlnLeuGln 532
OY 55 GAGGAATCCCAATTTGAGGAAAGCACTTACTGTAATATCAGAGAACTTGTGA 114
Db 593 GlnLeuThrGlnLeuGlnSerGluArgThrGlnLeuSerGlnValThrAspArgGlu 612
OY 115 GATCTTAAGAGCACTAAAGCATTAAGATTCTTGCTGCTCTACTTACTTGAACCT 174
Db 613 ArgLeuThrArgSpLeuGlnArgGlnGlnTyrGlnTyrGlnTyrGlnLeuAlaArg 632
OY 175 GTTGTGTGCTTGT-----TTGAATGCTCAGCATGAGCTGTTCTTTC 222
Db 633 AspValAlaLeuGlnAlaAlaGlnGlnGlnGlnAlaArgTyrHisAlaAlaValPhe--- 651
OY 223 CAATCCATTAATGTTCAATATGACAGACATC----- 255
Db 652 -----HisLeuGlnThrLeuLeuGlnLeuAlaThrGlnTyrGlnLeuAsp 665
OY 256 -----GAAAGACTGCTTAAGAAAGAGATGACTTGTCTGCA----- 294
Db 666 AlaLeuArgGlnArgAlaLeuAlaGlnArgSpGlnAlaAlaAlaGlnLeuAspAla 685
OY 295 -----CTAGTTCCGTAAGG 309
Db 686 AlaAlaSerThrSerGlnAsnAlaArgGlnSerAlaCysGlnArgLeuThrSerLeuGln 705
OY 310 AGCAGCTTGCGATACGAGCAAGAGAGCAAGAGCTTATGACAGGTGAAA----- 363
Db 706 GlnGlnLeuArgGlnSerGlnArgAlaAlaGlnGlnLeuAlaSerGlnLeuGlnAlaThr 725
OY 364 CAAGTTTGCATTAATCTGAGAGAGCCCAATTTGAAAAAACAAGCTTAAATCCAGTGT 423
Db 726 AlaAlaAlaLysSerSerAlaGlnGlnAspArgGlnAsnThrArgAlaThrLeuGln--- 744
OY 424 GACCACTTGAGGAGAGAGCTGAGGAGCGAGCGAGCACTTAAAAAAGAACTTGACATCT 483
Db 745 GlnGlnLeuArgGlnSerGlnAlaArgAlaAla-----GlnLeuAlaSer 759
OY 484 CAGCAAGAG-----AAAAGGCCATTGAGAAAGACATGATGAAAAAGGAA 528
Db 760 GlnLeuGlnAlaThrAlaAlaLysMetSerAlaGlnGlnAspArgGlnAsnThrArg 779
OY 529 ATPACGAAAGAAAGGAGATGATGAGTCAAAAGATGTTGCTTCAGAAATATGCC 588
Db 780 AlaThrLeuGlnGlnGlnLeuArgAspSerGln-----GlnArgAlaAla 794
OY 589 CAACGTGGAGCCAGGTGAGAAAGTTTACAAGAAAGATTCACCTATTAATCAACTG 648
Db 795 GlnLeuAlaSerGlnLeuGlnSerThrThrAlaAlaLysMetSerAlaGlnGlnAspArg 814

```

```

OY 649 GAGGAATTCAAAGCCAGCTGGCTTCGCGAAATGGATGTCAAAAGTGTGTGAGAA 708
Db 815 GlnSerThrArgAlaThrLeuGlnGlnGlnLeuArgSpSerGlnGlnArgAlaAlaGln 834
OY 709 ATGCCGTATCAGCTGAATTAACCAACATGGAGAGATGAGCAAAAGAGACACAGA 768
Db 835 LeuAlaSerGlnLeuGlnSerThrThrAlaAlaLysMetSerAlaGlnGlnAspArgGln 854
OY 769 GAGTTTCAGACCAAAACTAACAGGAGATCTGAATTAAGATCAGAAATAGAGAAATTG 828
Db 855 SerThrArgAlaThrLeuGlnGlnGlnLeuArgGlnSerGlnGlnAlaAlaGlnLeu 874
OY 829 AGAATAGAACTGTGATGAA-----AGCAAAACACACTTGAACAGAGAGACAGAG 879
Db 875 AlaSerGlnLeuGlnSerThrThrAlaAlaLysMetSerAlaGlnGlnAspArgGlnSer 894
OY 880 GCAGCCCTGGCCAGAGAGAGAGTGCCTGAGA---CTAACACAACTGCTGGCCGAATCGAG 936
Db 895 ThrArgAlaThrLeuGlnGlnGlnLeuArgSpSerGlnGlnArgAlaAlaGlnLeuAla 914
OY 937 CACCACTGCACCTCACCCAGATCTGAAATATGCTCAACTCAGTCAGTCAAGAAAAAGGTATACA 996
Db 915 SerGlnLeuGlnAlaThrAlaAlaLysMetSerAlaGlnGlnAspArgGlnAsnThr 934
OY 997 TATGATTAATTTGGGAAAGTTTACAGAGAGAAAGTAAGAA 1035
Db 935 ArgAlaAlaLeuGlnGlnGlnLeuArgSpSerGlnGln 947

```

Search completed: March 21, 2003, 13:07:07
Job time : 45.0233 secs